How complexity increases in development : An analysis of the spatial-temporal dynamics of Gene expression in Ciona intestinalis

Salvador-Martinez, Irepan

2017-04

Salvador-Martinez , I & Salazar-Ciudad , I 2017 , ' How complexity increases in development : An analysis of the spatial-temporal dynamics of Gene expression in Ciona intestinalis ' , Mechanisms of Development , vol. 144 , no. Part B , pp. 113-124 . https://doi.org/10.1016/j.mod.2017.02.001

http://hdl.handle.net/10138/232387
https://doi.org/10.1016/j.mod.2017.02.001

How Complexity Increases in Development: An Analysis of the Spatial-Temporal Dynamics of Gene Expression in *Ciona intestinalis*

Irepan Salvador-Martínez[1] and Isaac Salazar-Ciudad[1,2*]

[1] Evo-devo Helsinki community, Center of Excellence in Experimental Computational Developmental Biology, Institute of Biotechnology, University of Helsinki, Helsinki, Finland.

[2] Genomics, Bioinformatics and Evolution. Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain.

* Corresponding authors
Email: isaac.salazar@helsinki.fi (IS-C)
Telephone: +35 829 415 9892

Abstract

The increase in complexity in an embryo over developmental time is perhaps one of the most intuitive processes of animal development. It is also intuitive that the embryo becomes progressively compartmentalized over time and space. In spite of this intuitiveness, there are no systematic attempts to quantify how this occurs.

Here, we present a quantitative analysis of the compartmentalization and spatial complexity of *Ciona intestinalis* over developmental time by analyzing thousands of gene expression spatial patterns from the ANISEED database. We measure compartmentalization in two ways: as the relative volume of expression of genes and as the disparity in gene expression between body parts. We also use a measure of the curvature of each gene expression pattern in 3D space. These measures show a similar increase over time, with the most dramatic change occurring from the 112-cell stage to the early tailbud stage. Combined, these measures point to a global pattern of increase in complexity in the *Ciona* embryo. Finally, we cluster the different regions of the embryo depending on their gene expression similarity, within and between stages. Results from this clustering analysis, which partially correspond to known fate maps, provide a global quantitative overview about differentiation and compartmentalization between body parts at each developmental stage.

1. Introduction

The increase in complexity in an embryo over developmental time is perhaps one of the most intuitive processes of animal development. In spite of its intuitiveness, there is no consensus about how to define or measure this complexity.

A measure of morphological complexity that has been favored by many authors is the number of cell types that compose an organism (Bell and Mooers, 1997; Bonner, 2004; McShea, 1996). Using the number of cell types as a proxy of morphological complexity, it can be said that during metazoan development, complexity increases as the zygote divides and differentiates into an adult with multiple cell types. This simple definition of complexity has its complications, as there are no clear-cut consensus criteria about how to define a cell type or about when a new cell type has arisen during development. In addition, this definition does not take into account that embryos not only get more cell types, but these become organized in specific spatial patterns that seem to increase in "complexity" over developmental time.

This latter aspect of how complexity increases over development has been described as the progressive compartmentalization of the embryo in space (Carroll et al., 2001; Davidson, 2001). Carroll (Carroll et al., 2001) conceptualizes this at the level of gene expression. According to Carroll, the embryo becomes compartmentalized over time as genes progressively restrict their expression to smaller and smaller regions of the embryo during development. First, there are genes having a broad expression in the embryo that define the main axes of the body; then, field-specific selector genes defining compartments (for example organs and appendages); and finally, cell type specific selector genes expressed in specific cell types (e.g., muscle and neural cells). This process can involve both genes 1) whose expression correspond only to one of these categories or 2) that are first expressed widely and then restrict their expression in space (e.g. to cell types). In the latter case, the area of expression of a gene should decrease over time and, in addition, different regions of the embryo should express more different sets of genes as development progresses (they should have more disparity in gene expression).

Up to this point there has not been any systematic quantitative attempt to measure any aspect of compartmentalization. In other words, compartmentalization seems quite apparent when qualitatively comparing a small number of gene expression patterns in space, but is this still the case when many genes are considered using quantitative statistical methods?

Other developmental biologists (Gawantka et al., 1998; Struhl, 1991) understand complexity as the diversity of gene expression patterns in space and not directly as compartmentalization. In

that respect, little attention has been drawn to the shape of the gene expression domains themselves that is measured by the distribution in space of the cells expressing a given gene.

For example, an expression domain can have the shape of a simple circular spot, that of a stripe, or that of an irregular spot with a contour of any arbitrary shape. Complexity measures taking into account this irregularity, or roughness, have been devised to the analysis of animal morphology (Bunn, et al., 2011; Winchester, 2016).

Here, using data from the ANISEED database (Tassy et al., 2010; Brozovic et al., 2016) we analyzed the expression information of 2518 genes (S1 Table), expressed in at least one of six developmental stages of the ascidian *Ciona intestinalis* (*C. intestinalis*) (32-cell, 64-cell, 112-cell, early tailbud, mid tailbud and late tailbud stages), to measure quantitatively the spatial dynamics of gene expression through embryonic development. Our analysis is, thus, spatial and not only temporal as in many previous studies (Arbeitman et al., 2002; Bozinovic et al., 2011; Levin et al., 2012; Schep and Adryan, 2013). Our study takes a statistical developmental biology approach since it does not focus on a single gene, transduction pathway, gene network or organ but on a large number of genes for a number of stages and for the whole embryo. The aim is to quantitatively describe how gene expression changes in space over developmental time, and how this can be related to intuitive notions of complexity in the *Ciona* embryo. We measure compartmentalization as volume of expression, compartmentalization as disparity between tissues (the difference in which genes are expressed in different regions measured as 1 minus their gene expression correlation), and roughness of the spatial patterns of gene expression over the embryo.

We chose these measures because they are simple, relatively intuitive, and reflect existing ideas on how the embryo becomes more complex over time, as explained above. In this regard, these measures take into account different aspects of complexity in embryonic development. Our measures have been used before in a different species, *Drosophila melanogaster*, for the same questions (Salvador-Martínez and Salazar-Ciudad, 2015). Taking these three measures together, our analysis shows that, over time, the embryo becomes compartmentalized in distinct regions of gene expression that are smaller, more different between each other, and have a more rough shape. This general trend is shown in a higher or lower degree depending on the gene and whether than gene is a transcriptional factor, a growth factor or some other type of gene.

2. Methods

2.1. Expression data

In this article we use data from the ANISEED 2015 database (Tassy et al., 2010; Brozovic et

al., 2016). This database integrates expression data from large-scale *in situ* hybridization (ISH) studies with embryo anatomical data of ascidians (Tassy et al., 2010; Brozovic et al., 2016). The ANISEED 2015 includes 27,707 *C. intestinalis* gene expression profiles by ISH for about 4,500 genes acquired from more than 200 manually curated articles (Brozovic et al., 2016) including expression data from the Ghost database in *C. intestinalis* (Satou et al., 2005). The Ghost database contains spatial expression patterns of developmentally regulated genes derived from more than 6,000 whole mount ISH experiments (Fujiwara et al., 2002; Imai et al., 2004; Kusakabe et al., 2002; Ogasawara et al., 2002; Satou et al., 2001). The expression data is represented in the ANISEED database using an ontology-based anatomic description of the embryos and can be explored by users with a browsing tool (http://www.aniseed.cnrs.fr/ aniseed/experiment/find insitu). Taking advantage of the invariant ascidian cleavage pattern and well-described lineage analysis (Conklin, 1905; Nishida, 1987), the cDNA spatial expression of many genes have been described at the single cell level up to the early gastrula stage (Imai et al., 2004).

We downloaded the ish.zip file from the *C. intestinalis* download section of the ANISEED database on 28th of December 2015, which contains a single XML file with the expression data for many developmental stages in *C. intestinalis*. We extracted the information of the 32-cell, 64-cell, 112-cell, early tailbud, mid tailbud and late tailbud stages (reported as stages 6b, 8, 10, 19/20, 21/22 and 23/24/25, respectively). The expression data for the first three stages is at the cell level, whereas the data for the tailbud stages is at the tissues or specific regions of the embryo level. For further analyses, we used only expression data from experiments reported to have wild-type phenotype, “public” publication status, with *in situ* hybridization as experiment design, and whose probe was assigned to a Kyoto Hoya (KH) (Satou et al., 2008) gene model (Kyoto Hoya is the name of the current *C. intestinalis* genome assembly). All this information is included in the original XML file. We excluded data from experiments whose image characterization was reported as “not sure” or too broadly as “part of whole embryo”. In the case of redundant experiments (i.e., when more than one experiment had expression information of the same gene and same stage) one experiment was randomly selected.

For the relative volume, disparity, and roughness, we analyzed the genes separately by stage. For the relative volume and roughness, we analyzed only the genes with expression. The number of genes analyzed is n=745 for the 32-cell stage, n=758 for the 64-cell stage, n=809 for the 112-cell stage, n=1082 for the early tailbud, 1092 for the mid tailbud and 887 for the late tailbud stages. For disparity, we analyzed all genes with expression information in each stage separately (in this case genes that are annotated as without expression are also considered). The number of genes analyzed for disparity is n=1604 for the 32-cell stage, n=1630 for the 64-cell stage, n=1642 for the 112-cell

stage, n=1507 for the early tailbud, 1341 for the mid tailbud, and 1266 for the late tailbud stages. The number of genes analyzed for the synexpression territories analysis is mentioned in section 2.4.

2.2. Relative volume

For the first three stages, we calculated the volume using the biometry data of the ANISEED database. We used the volume of expression instead of the area of expression (as in Salvador-Martinez and Salazar-Ciudad, 2015) because we have 3D data. We downloaded the biometry.zip file from the downloads section and used the files "Late_32-cell_stage_Amira_High-Resolution_1.txt", 64-cell_stage_Amira_1.txt" and "Early_112-cell_stage_Amira_1.txt". Each file has a quantitative description of the geometry of individual blastomeres, including the volume of each blastomere in percentage of the whole embryo (Tassy et al., 2006). We also downloaded the 3D embryo models for these stages from the ANISEED database, available at a single-cell resolution (Fig. 1A). For the tailbud stages, we used a 3D model of *C. intestinalis* mid tailbud embryo anatomy (Nakamura et al., 2012). We downloaded the file "3DVMTE_THratio1.86.wrl" from http://chordate.bpni.bio.keio.ac.jp/3DVMTE/. This file contains the 3D anatomical model of a stage 22 mid-tailbud II embryo (Fig. 1B), which consists of a mesh of points and a mesh of polygons in 3D describing the external morphology of the embryo and the location of each tissue within it. We manually segmented each tissue into a separate 3D file (Fig. 1B). As a result, we obtained a 3D description of each tissue. We then processed them using filters of the Meshlab software version v1.3.3_64bit (Meshlab Visual Computing Lab ISTI-CNR). To estimate the volume of each tissue, we first modified the surface mesh for each tissue in the tailbud embryo by discarding all "internal" points (like inter-cellular contact surfaces). A common procedure to do this is to take random point samples over a 3D surface and then reconstruct a new surface from those points. We took each random points sample (creating a point cloud) by a Poisson-disk sampling algorithm (Corsini et al., 2012) (5,000-50,000 points depending on the tissue), then we reconstructed the surface using a Poisson surface-reconstruction algorithm (Kazhdan et al., 2006). Finally, the volume of each mesh was calculated from Meshlab intrinsic functions. For the genes that were expressed in several tissues, we repeated this procedure taking all the tissues where the gene is expressed and removing all the internal points (see Fig 1B for a gene that is expressed in the mesenchyme, palp, notochord, and muscle).

3D embryo models for early and late tailbud embryos were not available, so we used the 3D embryo model for the mid tailbud stage for all tailbud stages. In these stages the main morphogenetic process is the elongation by cell intercalation (Hotta et al., 2007), so the differences between these stages are largely restricted to tail length and width. As in the tailbud embryo 3D

model the Trunk lateral cells (TLC) were not distinguishable morphologically from the mesenchyme (Nakamura et al., 2012), we defined the volume of the TLC as 10% of the mesenchyme volume. This rough approximation was based on the number of cells of the mesenchyme (202 cells) and TLCs (16 cells) (Nakamura et al., 2012).

2.3. Roughness (Dirichlet Normal Energy or DNE)

Previously, we developed a measure called roughness (Salvador-Martínez and Salazar-Ciudad, 2015) that measures the curvature of the contour of a 2D gene expression pattern compared with the contour of a circle of the same perimeter. To provide a similar measure of curvature in 3D, we used the Dirichlet normal energy (DNE), which quantifies the deviation of a surface from being planar (Bunn et al., 2011). DNE values increase with both convexities and concavities in a surface. DNE is calculated as the sum of energy values *e(p)* across a polygonal mesh surface. A brief explanation of the method follows; for a detailed description of the method and its mathematical background, see (Bunn, et al., 2011; Winchester, 2016). The energy value *e(p)* quantifies change in the normal map around a polygonal face. The calculation of the energy value *e(p)* of a polygon can be exemplified with a schematic diagram as in Fig. 2B. First, the polygon is characterized by vectors *u* and *v*, which represent the edges of the polygon. Then, normal unit vectors are estimated as the normalized average of normal vectors of the triangle faces adjacent to each vertex (Fig. 2B). If vertex normals are translated to a common origin point, their end points form a polygon with edge vectors n_u and n_v, which represents the spreading of n_u and n_v (Fig. 2C). In a simplistic way, DNE can be defined as the spreading of n_u and n_v relative to the spreading of *u* and *v*. Polygons on more curved surfaces will produce greater relative spreading of n_u and n_v (Bunn et al., 2011; Winchester, 2016). More explicitly e(p)= tr(G^{-1} H), this is the trace of the produce of the matrices G^{-1} and H. These matrices are:

G = $\begin{pmatrix} \langle u,u \rangle & \langle u,v \rangle \\ \langle v,u \rangle & \langle v,v \rangle \end{pmatrix}$ and H = $\begin{pmatrix} \langle n_u,n_u \rangle & \langle n_u,n_v \rangle \\ \langle n_v,n_u \rangle & \langle n_v,n_v \rangle \end{pmatrix}$ and where < , > stands for dot product. Then the DNE for a surface is the sum of this energy over all faces weighted by the area of each surface.

$$\text{DNE} = \sum e(i) \cdot area(i)$$

To calculate the DNE, we used the Morphotester software version 1.1.2 (Winchester, 2016) from the webpage “http://morphotester.apotropa.com/”. The DNE was calculated using the “implicit

fair smooth” option, which reduces surface mesh noise that could disproportionately affect DNE values (Winchester, 2016). The 3D embryo models that we used (described above) allow to extract the 3D information of single cells so the reconstruction of any expression pattern in 3D can be done by creating a 3D model containing only the cells/tissues that constitute such expression pattern (as the expression information in ANISEED consist of a list of cells or tissues where the gene is expressed in a given stage). To calculate the DNE for each expression pattern, a new surface mesh (that is common to all the cells/tissues that form a gene expression) needs to be created, eliminating any internal surface (as in intercellular interfaces). Importantly, the DNE is invariant to the orientation or scaling of meshes, but is proportional to the number of polygons comprising a mesh (Bunn et al., 2011; Winchester, 2016), such that meshes with more polygonal faces will have higher DNE values than meshes of similar shape with less polygonal faces. To account for this, all gene expression patterns were reconstructed to have the same number of polygonal faces.

The surface reconstruction process described above was also done with the Meshlab software. For surface reconstruction, a common strategy is to first generate random sample points in the surface we want to reconstruct; then, a new surface is reconstructed using those sample points. We created 10,000 sample points generated with a “Poisson-disk Sampling” algorithm (Corsini et al., 2012). Then the surface is reconstructed using a “Poisson surface-reconstruction” (Kazhdan et al., 2006). As the reconstructed surface could have less polygons than desired, we subdivided the polygons with 3 iterations of the “Subdivision Surfaces: LS3 Loop" algorithm (Barthe and Kobbelt, 2004). To ensure that the original and the reconstructed mesh were most similar, vertex normal and geometry attributes of the former were transferred to the latter by the "Vertex Attribute Transfer" filter. Finally, the mesh was simplified (vertex number was reduced while preserving the overall shape of the 3D model) using a "Quadric Edge Collapse Decimation" algorithm (Garland, 1997) to produce a mesh of 10,000 polygonal faces. For the 1,000 and 100 polygonal faces, subsequent filters of "Quadric Edge Collapse Decimation" were applied.

2.4. Synexpression territory analysis

In our dataset, the number of genes with expression information throughout all the stages is low (n=270). Because of this, and because the information in the early stages is at the cell level whereas in the tailbud stages it is at the tissue level, we performed two separate synexpression territory analyses one with the 32-cell, 64-cell and 112-cell stages (n = 1550 genes), and another one with the early, mid and late tailbud stages (n = 820 genes). In both cases, we will refer to the embryo partitions (cell pairs in the early stages and single tissues in the tailbud stages) as regions.

For each region in a stage (there are 104 regions in the three early stages and 46 in the three

tailbud stages) we built a binary expression vector describing whether each gene was expressed or not. Each expression vector had 1550 components (genes) in the early stages and 820 in the tailbud stages, with binary values of the expression of every gene in that region (1 for expression and 0 for without expression). Then, we computed pairwise similarities between regions as the Pearson correlations between their expression vectors using the function *corSimMat* of the R package *apcluster* (Bodenhofer et al., 2011). With this function, two regions are similar if they are positively correlated. Using this similarity matrix, we performed a hierarchical clustering of regions using the function *hclust* of the R package *stats* version 3.11 (R Core Team, 2015) with the average method UPGMA (Unweighted Pair Group Method with Arithmetic Mean) and an Euclidean distance function. The resulting trees have as many terminal branches as there are regions over stages. We cut the resulting dendrogram in 24 and 10 clusters (or territories) to facilitate the analysis. We use the term "synexpression territory" to refer to any cluster of regions found in the dendrogram resulting from the previous analysis.

2.5. Cells with unique expression profile

We measured how many cells in the first three stages have a unique pattern of gene expression (i.e., a unique combination of genes expressed). Then, for each stage, we measured the number of blastomere pairs with a unique gene expression profile and their proportion in respect to half the number of blastomeres in each stage (as the *Ciona* embryo is bilaterally symmetrical).

2.6. Spatial disparity

Disparity in a given stage measures the overall difference in gene expression between regions. The pairwise similarity in gene expression between two regions is calculated as Pearson's correlation using the same expression vectors as before. As the disparity is a dissimilarity measure, the disparity between two regions is calculated as 1 minus their correlation, so if two regions would have exactly the same expression vectors, their correlation would be 1 and their disparity 0.

Disparity (i,j) = 1- corr(i,j) where i and j are two different cells or tissues.

Notice that given that the correlation ranges from -1 to 1, the disparity ranges from 0 to 2; therefore, disparity values between two regions greater than 1 would reflect that their gene expression is negatively correlated. The mean disparity in gene expression in a stage is the mean of all the pairwise disparities for all the regions. Disparity is a measure of gene expression complexity in space since high disparity implies that even close regions have quite different patterns of gene

expression.

2.7. Transcription factors

To check if members of the same transcription factor gene family have a mean relative volume or mean disparity that differs in a significant way from that of the rest of the genes, we used the comprehensive list of transcription factors (http://ghost.zool.kyoto-u.ac.jp/TF_KH.html) deposited in the Ghost database (last access in July 2015). This list is based mainly in Imai et al. (2004). Imai et al. (2004) determined the expression profiles of 389 transcription factors (TFs) and 118 signaling molecules (SIGs) genes from the egg to mid tailbud embryos. The transcription factor genes are divided into nine gene families: basic helix-loop-helix (bHLH), homeodomain (HD), Fox, ETS, bZIP, nuclear receptor (NR), HMG, T-box transcription factors, or as other (mainly with diverse Zinc finger genes). As the text-based annotation from Imai et al. (2004) is included in the expression patterns download file, we used these comprehensive lists as a reference to categorize the genes as transcription factors genes. The list of TFs used in here is shown in S2 Table.

2.8. Statistical analysis

For our three measures (relative volume, disparity, and roughness) we tested for significant differences between each stage and its subsequent stage with a Kruskal-Wallis rank sum test. To test whether the members of the TF families showed significant differences in their relative volume and disparity when compared with the rest of the genes, a permutation test was applied. This is, for each gene family we first measured, for both relative volume and disparity, the observed mean values of the two groups: one group labeled as the gene group of interest (e.g., “Homeobox genes”) and the other as the rest of the genes (e.g, “non-Homeobox genes”). Then, we shuffled labels 10,000 times, measuring each time the mean values for the randomly permuted groups. The p value is calculated as the fraction of times the permuted differences is greater or equal than the observed difference, out of the total number of permutations.

3. Results

3.1. The volume of expression of genes decreases over time

As described in the introduction, the progressive compartmentalization of the embryo should be reflected in a progressive decrease in the area of expression of genes. In Figure 3A, we show the volume of the domains of expression of genes relative to the whole embryo volume across developmental stages. The volume of expression is measured relative to the embryo volume in each

stage. For the first three stages, the volume of expression of a gene is simply the sum of the volume of the cells expressing it. This latter measure is very similar to the proportion of cells in which a gene is expressed (see Fig. S1). In the last three stages, however, we do not have information about gene expression at the single cell level but at the level of the tissues defined in ANISEED for those stages (see methods for details, Table 1 and Figure S11). The number of tissues in the tailbud stages (15; see Figure 7A and Table 1) is comparable to the number of cell pairs (as *Ciona* is bilaterally symmetric) in the earliest stage analyzed here, so in principle the relative volume could be as low or as high in the early stages as in the late stages. This allows us to roughly compare these volume measures over all the stages in spite of the different ways in which early and late stages are described. As we can see in Fig. 3A, the average relative volume of expression decreases over developmental stages with the strongest decrease occurring after the initiation of gastrulation (in the transition between the 112-cell stage and the early tailbud stage). There are some genes, however, that are restricted to a small part of the embryo from very early on, shown in Fig. 3A as the Min values (the 10% of genes that have the lowest volume of expression).

3.2. Transcription factors compartmentalization

We analyzed if this decrease in gene expression volume was different for transcription factors (TFs) compared to the rest of genes, as these genes are considered to have a major role in driving the compartmentalization of the embryo. TFs decrease their average volume of expression faster than non-TFs. The distribution of these volumes of expression in the early stages is, however, very asymmetric with most genes expressed ubiquitously. When comparing between TFs and non-TFs the 10% of genes with the lowest volume of expression, it is still the case that TFs show smaller volumes of expression than non-TFs, at least from 64-cell stage until late tailbud stage (Fig, 3B). TFs, thus, are compartmentalized earlier (i.e., expressed in a smaller volume of expression) than other genes.

To explore if our results can be attributed to maternal or zygotic TF genes, we repeated our statistical analysis for each category (comparing them with non-TF genes of the same maternal versus zygotic categories). The classification of a gene as maternal or zygotic was extracted from Matsuoka et al. (2013). We found no differences in volume between maternal and zygotic genes at any developmental stage (Fig. S2).

We also compared TFs between families, using the categories of TFs families used by Imai (Imai et al., 2004). We analyzed whether the members of these families showed significant differences in their mean relative volume when compared with the rest of the genes (permutation test, 10000 permutations). As shown in Fig. S3, only the T-box gene family shows a significantly

lower relative volume in all the early stages. At the 32-cell stage, when six T-box genes are expressed, four of them have a restricted gene expression. Six TF families show a significantly lower relative volume at the 112-cell stage: BZIP, T-box, bHLH, HMG, Nuclear Receptor, and Other-TFs. We found that none of the main TF families showed a lower volume of expression than the rest of genes in the tailbud stages. Only the category of “Other-TFs” showed a higher volume of expression in such stages. The majority of these genes are deposited maternally and additionally, more than 50% of the maternal TFs of this category are expressed ubiquitously until the mid tailbud stage (see Fig. S4).

It could be that an under or over-representation of TF genes in our sample would affect our results. The ANISEED database is composed of data coming from many ISH experiments, including experiments specifically aiming at TF genes (Imai et al., 2004; Miwata et al., 2006). These studies performed a systematic description of TF gene expression profiles during embryogenesis, describing the expression of more than 80% of *C. intestinalis* TFs. Based on this, we would expect these genes to be overrepresented in the ANISEED database and therefore, in this analysis. Indeed, in our analysis, TFs represent 9.7% of the genes analyzed, whereas the proportion of TFs in the *C. intestinalis* genome is estimated to be around 4% (Dehal, et al., 2002). In this sense, the over-representation of TF genes could in principle affect the global pattern we observe in compartmentalization, especially in the 64-cell and 112-cell stages, as we have shown that in those stages TFs are expressed in a restricted manner. To test if the global compartmentalization pattern could be affected by this over-representation, we repeated the analysis of relative volume excluding TFs and tested for significant differences between successive stages (Fig. S5). The statistical test of this analysis shows the same as the test including those genes; from the 32-cell to the early tailbud stage, the relative volume of expression becomes more restricted every stage. Thus, our main conclusions on gene compartmentalization are not affected by the over-representation of TFs.

3.3. Spatial disparity

In principle, it cannot be completely ruled out that genes decrease their volume of expression but that the embryo does not become more compartmentalized. This would be the case if the genes with restricted expression would be expressed in the same set of cells/tissues in the embryo. To evaluate that, we calculated how similar are the list of genes that are expressed in any two cells (for early stages) or tissues (for late stages) in each of the stages. This wass done through Pearson's correlation, 1 minus this correlation is what we call the disparity between any two cells or tissues (see methods for details). Notice that in this case, early stages are not directly comparable to late stages since the latter are analyzed at the level of tissues made of several cells. In these latter

stages, thus, we can only analyze the genes that are expressed in at least one cell and, in principle, the more cells there are in a territory, the larger the number of genes that can be differentially expressed. The early stages are, however, comparable between themselves and the late stages are also comparable between themselves. For each stage we calculated the average disparity (see Fig. 4A). Figure 4A indicates that disparity increases between early stages and also between late stages, at least between the 64 and 112-cell stages, and between the early and mid tailbud stages. This indicates not only that gene expression becomes progressively restricted to smaller regions of the embryo (smaller volume of expression shown in Fig. 3), but that, in addition, these regions express progressively more different combinations of genes. In other words, the embryo becomes compartmentalized in regions that are progressively smaller and more different from each other.

3.4. Transcription factors have a higher disparity during earlier developmental stages

As with the relative area, we analyzed if the disparity in gene expression area was different for TFs compared to the rest of genes. TFs showed significantly greater disparity than the other genes in the 32-cell and 64-cell stages (Fig. 4B). Conversely, in the early and mid tailbud stages the TFs have significantly lower disparity than the rest of the genes, which again may be driven by the genes in the "Other-TFs" category (Fig. S6). Four TF families showed a significant greater disparity than the rest of genes in the 64-cell stage: T-box, Ets, HMG, and Other-TFs. The HMG family showed a higher disparity than the rest of genes in all the three early stages, and no TF family showed a significantly higher disparity in the tailbud stage (Fig. S6).

3.5. Roughness increases over developmental stages

As described in the introduction, the measures of volume and disparity tell nothing about the shape of the territory of gene expression in space. One possibility is that as development progresses these territories progressively transform from more or less spherical shapes to more complicated or rough shapes, as seems to occur at the level of morphology. We chose to use one such measure of roughness that has been used for 3D morphology (Winchester, 2016). The Dirichlet Normal Energy or DNE is a measure of complexity that considers the overall curvature of the 3D shape of a surface (in our case the 3D spatial domain of expression of a gene). This measure is roughly equivalent to the sum of the curvature, or roughness, at each point of a surface (see methods for detail). Relative volume and DNE are independent measures. In fact, the DNE is known to be invariant to volume (Winchester, 2016); a territory that has a small volume does not imply that the DNE is small, and there could then be, in principle, embryos with high average DNE and low volume of expression or *vice versa*. The DNE and disparity are not necessarily correlated either; this can be illustrated with a

simple example. If in a blastula a large proportion of the genes are ubiquitously expressed and a small proportion are expressed in a single cell, both roughness and disparity would be low. However, if a small proportion of the genes are expressed in all the embryo and a large proportion in single cells, the disparity would be high but the roughness would be the same (or very similar) than in the previous case (see Fig. S7 and S8 for a visual depiction of the relationship between these measures).

We chose to measure the DNE at different spatial scales for each gene expression pattern. This allows us to take into account that while some patterns may be very rough at a low spatial scale, others may be rough at a high spatial scale. For example, in the case of a pattern with a single cell represented with many polygons (e.g., 1,000), the DNE would measure the curvature at a sub-cellular level, so an epithelial cell will tend to have lower curvature than a fibroblast, as the former's surface is more regular. At a “low resolution” (less polygons), the DNE would measure the curvature at a multi-cellular level (see Fig. 5), like the bending of an epithelium.

Our results (Fig. 5) show that roughness increases with developmental time. The major increase is between the 112-cell and the tailbud stages. The Max values (mean value of the 10% of genes that have the highest roughness) show also a substantial increase already between the 64-cell and 112-cell stages (in the 100 and 1,000 polygonal faces). These Max values are informative about the overall morphological complexity of the embryo in a given stage. Median values are not very informative about morphological complexity since even in very complex organisms there are always genes that need to be expressed in most cells (for example genes of core metabolism that all cells need). In fact, we see that the Min values (mean of the first decile) remain practically constant during development, showing that there is always a proportion of genes with a very simple expression pattern. In the analysis of 100 and 1,000 polygons per mesh, the mean DNE increases significantly in the 112-cell and early tailbud stages transition from its previous stage (one-way ANOVA pvals < 0.001), suggesting that it is during these transitions that the roughness of gene expression increases the most. Interestingly, in the 1000 and 10000 polygons per mesh analysis there is also an increase in the mean DNE of the late tailbud stage in respect to the mid tailbud stage (one-way ANOVA pvals < 0.05).

Note that the fact that we extract gene expression differently in the early and late stages may introduce a bias in roughness. Because during post-gastrula stages ANISEED describes the embryo in multicellular regions (as opposed to individual cells in pre-gastrula or gastrula stages), expression roughness within these regions can not be detected. This can potentially lead to an underestimation of roughness in the tailbud stages. We find that roughness is larger in the tailbud stages and, therefore, this potential underestimation does not affect our finding that later stages have higher

roughness.

3.6. Spatio-temporal dynamics in the *Ciona* embryo

To better describe how much the different parts of the embryo differentiate from each other over time, we measured the gene expression similarity between regions and organized them into a dendrogram to explore the hierarchy of similarity between them. In the first three stages, these regions are individual cells (in each side of the bilaterally symmetric embryo), whereas for the tailbud stages these are tissues. Because of this difference, we built a dendrogram for the three first stages and another for the three last stages. Thus, every terminal branch corresponds to an individual cell in the first stages and to a region in the tailbud stages. These two dendrograms were produced by means of a hierarchical clustering algorithm (see methods) (Fig. 6A and Fig. 7A). Each of these sets of same-branch regions is called a synexpression territory or just a territory.

The dendrograms are analogous to a phylogenetic tree between species. Instead of species, however, we have cells or tissues, and the distances between those are based on the proportion of genes they have in common. In contrast to phylogenetic trees, these dendrograms do not directly inform about cell genealogy or lineage (that in *C. intestinalis* is anyway well known), but about how different in gene expression are the different cells and regions in the embryo. This is also different from embryo fate maps, in the sense that a fate map informs about which cells, or embryo regions, give rise to which specific cell type or tissue, but it does not tell how different are those cells quantitatively (i.e., which proportion of genes they express differently) at a given stage. In addition, it can occur, as we find in a small number of cases in *Ciona*, that some groups of cells are more similar to other cells with a different fate (at the quantitative level of which genes they express) than to cells with the same fate. This can occur because fate can be disproportionally affected or determined by a small number of genes, (e.g., a small number of specific TFs). Since these genes determine fate, they are expected to regulate the expression of many other genes, but these regulatory consequences are not necessarily immediate, and then the majority of changes in gene expression at a given stage may not always be directly related to fate. Note that our measures of disparity and dendrograms reflect quantitative differences at the bulk level of all expressed genes and can not tell which of these changes would have important consequences for fate. We provide examples of that in the following paragraphs.

If we consider the general structure of one of such dendrograms (e.g, the number of major branches and the territories that comprise them), two extreme scenarios are possible *a priori*: 1) Territories cluster with other territories in the same developmental stage. In this scenario the majority of genes change their expression in a similar way over time independently of the region of

the embryo where they are expressed. 2) Territories in different stages cluster together. In this scenario, the clustered territories would be most probably related by cell lineage (in the case of two clustered territories the territory in the earliest stage would be the precursor of the territory in a later stage). This would reflect that each embryo part or region has gene expression dynamics that are largely unrelated to those of other parts or regions of the embryo. The early stages dendrogram follows scenario 1, as the territories cluster by stage. Thus, even if at the first three stages a high proportion of blastomeres express a nearly unique combination of transcriptional factors (Imai et al., 2006; see next section), the bulk change in gene expression is common among all blastomeres.

Within each stage, most of the cells with the same fate branch together (Fig 6A, Fig. S9). There are, however, some exceptions that we describe below. In the 64-cell stage, the cell pairs B7.4 and B7.5 branch together with cell pairs A7.4, A7.8 (dark green synexpression territory in the 64-cell stage; Fig. 6B). The former are muscle progenitors, whereas the latter are nerve cord progenitors (Nishida, 1987; Nicol and Meinertzhagen, 1988a). The cells where these inconsistencies are found share the expression of many genes and have a high concentration of mitochondria (Fujiwara et al., 2002). Because of this, these genes have been reported to have a “mitochondria-like distribution” (Fujiwara et al., 2002). Since the dendrogram captures bulk similarity in gene expression, the shared expression of “mitochondria-like” genes might make these cells to branch together.

In the 112-cell stage, there are also dendrogram territories that are comprised of cells with different fates. One such territory (light green synexpression territory in the animal view of the 112-cell stage; Fig. 6B) includes neural plate (cell pairs a8.18, a8.20, and a8.26) and epidermis precursors (cell pairs a8.27, a8.28 and b8.20). This again might be explained by the genes with “mitochondria-like” distribution, as there are 21 genes in this stage with such expression pattern, that are expressed in those neural plate and epidermis progenitors. An example of a gene showing this expression pattern is the Ci-tubulin beta-02 gene (KH2012:KH.L116.85).

In the tailbud stages’ dendrogram (Fig. 7A), all but one branches correspond to the same tissue/cell type. The only territory that has more than one fate is the one that contains the notochord and endodermal strand from the early tailbud stage. This is also reflected by the higher proportion of shared expressed genes between these tissues (44%), than the proportion of genes shared between the endodermal strand (38.2%) or notochord (39.7%) of the early and mid tailbud stages (data not shown). This result indicates that in the early tailbud embryo, most tissues are already quite different at the level of gene expression, as they are closer to their descendant territories than to any other territory at the same stage (with the exception noted above). This is largely consistent with the bulk of other studies analyzing these stages at the level of individual or small sets of genes

(Corbo et al., 1997a; Di Gregorio and Levine, 1999).

3.7. Gene expression dynamics tracing cell-lineage relationships

We also explored the gene expression similarity between mother and daughter cells and between sister cells along lineages in development. Figure 8A shows the number of genes differentially expressed between each mother and daughter and between each pair of sisters (see S3 Table to see which are those genes). Fig. S10 shows, instead, the proportion of genes that are differentially expressed between mother and daughter and between sisters. The results in Fig 8A and S10 indicate that in the cleavage stages, every cell is more similar to its sister cell than to its mother or daughter cells, but that mothers and daughters are anyway quite similar. This should not be necessarily the case. It could be, for example, that after one cell divides, each daughter cell expresses a different set of 10 genes that were not expressed in the mother cell. In this case, each daughter cell would have 10 genes expressed differentially from its mother cell and 20 genes expressed differently between each other. Therefore, the difference in gene expression is higher between mother-daughter cells, even when the majority of daughter-cell differences are of less than 40 genes, a small subset of the genes we are considering.

Also, as can be seen in Fig. 8B, the proportion of blastomeres that have a unique pattern of gene expression decreases over time. The proportion of blastomere pairs with a unique expression pattern decreases from >80% in the 32-cell to <60% in the112-cell stage. It is interesting to see that, even when the number of cells with different gene expression increases in the cleavage stages, its proportion relative to the total number of cell decreases. This can be interpreted as to mean that the embryo is most compartmentalized in the early cleavage stages (32-cell stage in our analysis). However, as our disparity analysis shows, the proportion of the genes expressed differently is very low. So what happens in later stages is that the proportion of genes expressed differentially increases but the cells expressing totally different genes decreases (and the number of cells and of gene expression territories increases). This almost complete genetic partitioning of the ascidian early embryo has been noticed before (Imai et al., 2006). The earliest pattern formation in ascidians (which although have been studied mostly using *Halocynthia roretzi* as model organism they are thought to also apply to *C. intestinalis*) happen inside an oocyte that is already heterogeneous (at least with an animal pole and a vegetal pole with different maternal mRNAs), followed by cortical and cytoplasmic reorganizations between fertilization and first cleavage (e.g, cortical contraction towards the vegetal pole immediately after fertilization and dorsal-pole position caused by cyto-cortex translocation influenced by the spermatozoa entry point) combined with short-range signaling between individual cells (reviewed in: Lemaire, 2009; Sardet et al., 2007). This short-

range signaling, however, does not seem to partition each new cell that arises as the embryo keeps growing in cell number. This measure of compartmentalization is not suitable for the last three stages because in them, each region is made of several cells.

4. Discussion

The present analysis provides quantitative support for the view that complexity increases over developmental time in *C. intestinalis*. This holds for the different aspects of complexity that are captured by the relative volume of expression, roughness, and spatial disparity measures. According to the relative volume of expression, each gene is expressed in progressively smaller regions in the embryo until the early tailbud stage. The disparity reveals that different regions (cells or regions depending on the stage) of the embryo express increasingly different combinations of genes until the mid tailbud stage, that is, one stage after any significant change in expression volume. Lastly, our roughness measure (measured with the DNE) indicates that the complexity of the distribution in 3D space of the cells and tissues expressing a gene increases through development. Taking these three measures together, our results show that, over time, the embryo becomes compartmentalized in distinct regions of gene expression that are smaller, more different between each other, and have a more rough shape.

It is important to notice that our analysis is quantitative but merely descriptive. We can not say from our analysis what produces the observed changes in volume, disparity and roughness. These can be due to signaling between cells and tissues that change where genes are expressed, to morphogenetic events that change the spatial location of the cells expressing the different genes (and thus potentially the volume and roughness of expression of many genes), or to a combination of both. In fact, morphogenetic events modifying the shape of the whole embryo are likely to affect, at least, the roughness of some genes' expression while local morphogenetic events would affect the roughness of other genes' expression. These are all changes to be considered and that account for measuring complexity over time, irrespective of the underlying mechanisms.

Our approach also allows us to analyze the hierarchy of gene expression similarity between different regions of the embryo (between and within stages). We found that the gene expression of each cell in the early stages (32-cell, 64-cell and 112-cell stages) is more similar to the gene expression of any other cell of the same stage than to any cell related by lineage to it. This indicates that, even if in the early stages, a high proportion of blastomeres express a nearly unique combination of transcriptional factors (Imai et al., 2006; see section 3.9), the bulk of the change in gene expression over time is common among all blastomeres.

As expected, we found that the sub-division of the early embryo stages in synexpression

territories, produced by clustering blastomeres given a gene expression similarity threshold, largely coincide with the known division of the embryo based on fate maps (Nishida, 1987). This result coincides with the acknowledged early specification of the Ascidian embryo (Kumano and Nishida, 2007; Nishida, 1987; Satoh, 2011).

Our analysis detects that TFs have a more spatially restricted expression than other genes. This is consistent with their alleged leading role in driving pattern formation and the resulting compartmentalization of the embryo (Carroll et al., 2001; Davidson, 2001). Specific TF gene families that showed high disparity or low relative volume of expression (Fig. S2 and Fig. S6) have been already reported to have a crucial role in early development. T-box family genes, for example, show a low relative volume in all the early stages (Fig. S2). These genes, conserved in metazoan and several non-metazoan lineages (Sebé-Pedrós et al., 2013), are known for their important role in early cell fate specification in diverse metazoan species (reviewed in: Papaioannou, 2014; Showell et al., 2004). In *Ciona*, examples of these are Tbx6 and *brachyury*, crucial for muscle tissue formation (Yagi et al., 2005) and for notochord specification (Corbo et al., 1997b; Takahashi et al., 1999), respectively.

Part of our analysis is similar to the analysis made by Imai et al (Imai et al., 2006). They performed a hierarchical clustering of blastomeres in the 16-cell, 32-cell, 64-cell, and 112-cell stages, based on the expression profile of 53 zygotically expressed transcription factor genes. Our analysis is different from Imai et al, in two aspects: 1) we performed the hierarchical clustering using the blastomeres of different stages and 2) our analysis is not restricted to zygotic TF genes (we used both maternal and zygotic genes including but not restricted to transcription factors). These differences allow for different and complementary interpretations. First, including different stages into the clustering is informative of the overall differentiation process and can be used to discern between the various differentiation scenarios occurring in each stage transition (see above for a detailed description). Second, the reported number of “blastomere identities” by Imai et al. (blastomeres with specific combinations of zygotically transcribed TF genes), differs from the number of cells with a unique expression pattern reported here: for example, for the 32-cell stage, Imai et al., report 7 blastomere identities, whereas we found 13 blastomere pairs with unique expression profile.

In a previous study (Salvador-Martínez and Salazar-Ciudad, 2015), we performed a similar analysis in *Drosophila melanogaster.* Despite some differences between the databases, the main result that complexity increases over developmental time is consistent between the two species.
Nevertheless, there are some differences between these species. In *D. melanogaster* the major decrease in the global relative area of expression occurs previous to, or around gastrulation, whereas

in *C. intestinalis* the major decrease in the relative volume of expression occurs after gastrulation. This earlier compartmentalization in *D. melanogaster* could be related to its derived early development, namely, the syncytial blastoderm. During the syncytial blastoderm stage, the approximately 4,000 cell nuclei can “communicate” with each other only by TF genes (Jaeger, 2011). This rapid and highly dynamic process, facilitated by the direct cross regulation of gene expression, may be responsible for the early spatial restriction of a great proportion of developmental genes. In *Ciona*, early embryonic patterning is based on maternal determinants and signaling events usually between neighboring cells (Lemaire, 2009). These short-range signaling events that can act in a combinatorial way (Hudson et al., 2007) establish a unique TF combination in more than half of the blastomere pairs before gastrulation (Imai et al., 2006). Therefore, even when in *Ciona* many cell fates are already determined at these early stages and the embryo can be said to be already highly partitioned or compartmentalized, this is not evident at the global level of gene expression.

The results presented in this study, based on a statistical approach, provide a different and complementary view to the prevailing "individualistic" approach in developmental biology (Davidson, 2009), by which the role of a single gene (or of a small set of genes) in the development of a specific structure is investigated. A general description of the overall gene expression patterns can be acquired by comparing these individualistic gene studies (Kumano and Nishida , 2007; Lemaire, 2009). In our statistical approach, instead, we measure these patterns directly. The increase in complexity of an embryo (regardless of how complexity is defined) is perhaps the most apparent and characteristic phenomenon during development. In here, we have measured, in different ways, the dynamics of the complexity increase in the *Ciona* embryo.

Acknowledgments:

We thank J. Jernvall, O. Shimmi, M. Fortelius, M. Marín-Riera, M. Brun-Usan, P. Hagolani, R. Zimm, J. Moustakas-Verho, and the Evo-devo Helsinki community for comments. This research was funded by the Finnish Academy (WBS 1250271).

References

Arbeitman, M.N., Furlong, E.E.M., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., White, K.P., 2002. Gene expression during the life cycle of Drosophila melanogaster. Science 297, 2270–5. doi:10.1126/science.1072152

Barthe, L., Kobbelt, L., 2004. Subdivision scheme tuning around extraordinary vertices. Comput. Aided Geom. Des. 21, 561–583. doi:10.1016/j.cagd.2004.04.003

Bell, G., Mooers, A., 1997. Size and complexity among multicellular organisms. Biol. J. Linn. Soc. 60, 345–363.

Bodenhofer, U., Kothmeier, A., Hochreiter, S., 2011. APCluster: an R package for affinity propagation clustering. Bioinformatics 27, 2463–4. doi:10.1093/bioinformatics/btr406

Bonner, J.T., 2004. Perspective: The Size-Complexity Rule. Evolution (N. Y). 58, 1883–1890. doi:10.2307/3449441

Bozinovic, G., Sit, T.L., Hinton, D.E., Oleksiak, M.F., 2011. Gene expression throughout a vertebrate's embryogenesis. BMC Genomics 12, 132. doi:10.1186/1471-2164-12-132

Brozovic, M., Martin, C., Dantec, C., Dauga, D., Mendez, M., Simion, P., Percher, M., Laporte, B., Scornavacca, C., Di Gregorio, A., Fujiwara, S., Gineste, M., Lowe, E.K., Piette, J., Racioppi, C., Ristoratore, F., Sasakura, Y., Takatori, N., Brown, T.C., Delsuc, F., Douzery, E., Gissi, C., McDougall, A., Nishida, H., Sawada, H., Swalla, B.J., Yasuo, H., Lemaire, P., 2016. ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Res. 44, D808–D818. doi:10.1093/nar/gkv966

Bunn, J.M., Boyer, D.M., Lipman, Y., St Clair, E.M., Jernvall, J., Daubechies, I., 2011. Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination. Am. J. Phys. Anthropol. 145, 247–61. doi:10.1002/ajpa.21489

Carroll, S.B., 2001. Chance and necessity: the evolution of morphological complexity and diversity. Nature 409, 1102–9. doi:10.1038/35059227

Cole, A.G., Meinertzhagen, I.A., 2004. The central nervous system of the ascidian larva: mitotic history of cells forming the neural tube in late embryonic Ciona intestinalis. Dev. Biol. 271, 239–62. doi:10.1016/j.ydbio.2004.04.001

Conklin, E.G., 1905. The organization and cell-lineage of the ascidian egg. J. Acad. Nat. Sci. Philadelphia 13, 1–119.

Corbo, J., Erives, A., Di Gregorio, A., Chang, A., Levine, M., 1997a. Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. Development 124, 2335–2344.

Corbo, J.C., Levine, M., Zeller, R.W., 1997b. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124, 589–602.

Corsini, M., Cignoni, P., Scopigno, R., 2012. Efficient and flexible sampling with blue noise properties of triangular meshes. IEEE Trans. Vis. Comput. Graph. 18, 914–24. doi:10.1109/TVCG.2012.34

Davidson, E.H., 2001. Genomic Regulatory Systems: In Development and Evolution. Academic Press.

Davidson, E.H., 2009. Developmental biology at the systems level. Biochim. Biophys. Acta - Gene Regul. Mech. 1789, 248–249. doi:10.1016/j.bbagrm.2008.11.001

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., Harafuji, N., Hastings, K.E.M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I.A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-Bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D.N., Haga, S., Hayashi, H., Hino, K., Imai, K.S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K.W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M.M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P.D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., Rokhsar, D.S., 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298, 2157–67. doi:10.1126/science.1080049

Di Gregorio, A., Levine, M., 1999. Regulation of Ci-tropomyosin-like, a Brachyury target gene in the ascidian, Ciona intestinalis. Development 126, 5599–5609.

Fujiwara, S., Maeda, Y., Shin-I, T., Kohara, Y., Takatori, N., Satou, Y., Satoh, N., 2002. Gene expression profiles in Ciona intestinalis cleavage-stage embryos. Mech. Dev. 112, 115–27.

Garland, M., Heckbert, P.S., 1997. Surface simplification using quadric error metrics, in: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH ’97. ACM Press, New York, New York, USA, pp. 209–216. doi:10.1145/258734.258849

Gawantka, V., Pollet, N., Delius, H., Vingron, M., Pfister, R., Nitsch, R., Blumenstock, C., Niehrs, C., 1998. Gene expression screening in Xenopus identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. Mech. Dev. 77, 95–141.

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T., Ikeo, K., 2007. A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. Dev. Dyn. 236, 1790–805. doi:10.1002/dvdy.21188

Hudson, C., Lotito, S., Yasuo, H., 2007. Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of

distinct cell identities in the ascidian neural plate. Development 134, 3527–3537. doi:10.1242/dev.002352

Hudson, C., Yasuo, H., 2006. A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in Ciona embryos. Development 133, 2855–64. doi:10.1242/dev.02466

Hudson, C., Yasuo, H., 2005. Patterning across the ascidian neural plate by lateral Nodal signalling sources. Development 132, 1199–210. doi:10.1242/dev.01688

Imai, K.S., Hino, K., Yagi, K., Satoh, N., Satou, Y., 2004. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131, 4047–58. doi:10.1242/dev.01270

Imai, K.S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science 312, 1183–7. doi:10.1126/science.1123404

Imai, K.S., Satoh, N., Satou, Y., 2002. Early embryonic expression of FGF4/6/9 gene and its role in the induction of mesenchyme and notochord in Ciona savignyi embryos. Development 129, 1729–1738.

Jaeger, J., 2011. The gap gene network. Cell. Mol. Life Sci. 68, 243–74. doi:10.1007/s00018-010-0536-y

Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction, in: Polthier, K., Sheffer, A. (Eds.), Eurographics Symposium on Geometry Processing. Eurographics Association, pp. 61–70.

Kumano, G., Nishida, H., 2007. Ascidian embryonic development: an emerging model system for the study of cell fate specification in chordates. Dev. Dyn. 236, 1732–47. doi:10.1002/dvdy.21108

Kusakabe, T., Yoshida, R., Kawakami, I., Kusakabe, R., Mochizuki, Y., Yamada, L., Shin-i, T., Kohara, Y., Satoh, N., Tsuda, M., Satou, Y., 2002. Gene expression profiles in tadpole larvae of Ciona intestinalis. Dev. Biol. 242, 188–203. doi:10.1006/dbio.2002.0538

Lemaire, P., 2009. Unfolding a chordate developmental program, one cell at a time: invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. Dev. Biol. 332, 48–60. doi:10.1016/j.ydbio.2009.05.540

Levin, M., Hashimshony, T., Wagner, F., Yanai, I., 2012. Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. Dev. Cell 22, 1101–1108. doi:10.1016/j.devcel.2012.04.004

Matsuoka, T., Ikeda, T., Fujimaki, K., Satou, Y., 2013. Transcriptome dynamics in early embryos of the ascidian, Ciona intestinalis. Dev. Biol. 384, 375–85. doi:10.1016/j.ydbio.2013.10.003

McShea, D.W., 1996. Perspective: Metazoan Complexity and Evolution: Is There a Trend?

Evolution (N. Y). 50, 477. doi:10.2307/2410824

MeshLab Visual Computing Lab - ISTI - CNR http://meshlab.sourceforge.net/

Miwata, K., Chiba, T., Horii, R., Yamada, L., Kubo, A., Miyamura, D., Satoh, N., Satou, Y., 2006. Systematic analysis of embryonic expression profiles of zinc finger genes in Ciona intestinalis. Dev. Biol. 292, 546–54. doi:10.1016/j.ydbio.2006.01.024

Nakamura, M.J., Terai, J., Okubo, R., Hotta, K., Oka, K., 2012. Three-dimensional anatomy of the Ciona intestinalis tailbud embryo at single-cell resolution. Dev. Biol. 372, 274–84. doi:10.1016/j.ydbio.2012.09.007

Nicol, D., Meinertzhagen, I.A., 1988a. Development of the central nervous system of the larva of the ascidian, Ciona intestinalis L. I. The early lineages of the neural plate. Dev. Biol. 130, 721–36.

Nicol, D., Meinertzhagen, I.A., 1988b. Development of the central nervous system of the larva of the ascidian, Ciona intestinalis L. II. Neural plate morphogenesis and cell lineages during neurulation. Dev. Biol. 130, 737–66.

Nishida, H., 1987. Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. Dev. Biol. 121, 526–41.

Ogasawara, M., Sasaki, A., Metoki, H., Shin-i, T., Kohara, Y., Satoh, N., Satou, Y., 2002. Gene expression profiles in young adult Ciona intestinalis. Dev. Genes Evol. 212, 173–185. doi:10.1007/s00427-002-0230-7

Papaioannou, V.E., 2014. The T-box gene family: emerging roles in development, stem cells and cancer. Development 141, 3819–33. doi:10.1242/dev.104471

R Core Team, 2015. R: A Language and Environment for Statistical Computing.

Salvador-Martínez, I., Salazar-Ciudad, I., 2015. How complexity increases in development: An analysis of the spatial–temporal dynamics of 1218 genes in Drosophila melanogaster. Dev. Biol. 405, 328–339. doi:10.1016/j.ydbio.2015.07.003

Sardet, C., Paix, A., Prodon, F., Dru, P., Chenevert, J., 2007. From oocyte to 16-cell stage: cytoplasmic and cortical reorganizations that pattern the ascidian embryo. Dev. Dyn. 236, 1716–31. doi:10.1002/dvdy.21136

Satoh, N., 2011. Tunicate Embryos and Cell Specification, in: eLS. John Wiley & Sons, Ltd. doi:10.1002/9780470015902.a0001514.pub2

Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., Hamaguchi, M., Ishikawa, H., Chiba, S., Imai, K., Kano, S., Murakami, S.D., Nakayama, A., Nishino, A., Sasakura, Y., Satoh, G., Shimotori, T., Shin-i, T., Shoguchi, E., Suzuki, M.M., Takada, N., Utsumi, N., Yoshida, N., Saiga, H., Kohara, Y., Satoh, N., 2001. Gene expression profiles in Ciona intestinalis tailbud embryos. Development 128, 2893–2904.

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., Satoh, N., 2005. An integrated database of

the ascidian, Ciona intestinalis: towards functional genomics. Zoolog. Sci. 22, 837–43.

Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J., Wasserscheid, J., Dewar, K., Wiley, G.B., Macmil, S.L., Roe, B.A., Zeller, R.W., Hastings, K.E.M., Lemaire, P., Lindquist, E., Endo, T., Hotta, K., Inaba, K., 2008. Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol. 9, R152. doi:10.1186/gb-2008-9-10-r152

Schep, A.N., Adryan, B., 2013. A comparative analysis of transcription factor expression during metazoan embryonic development. PLoS One 8, e66826. doi:10.1371/journal.pone.0066826

Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gómez-Skarmeta, J.L., Ruiz-Trillo, I., 2013. Early evolution of the T-box transcription factor family. Proc. Natl. Acad. Sci. U. S. A. 110, 16050–5. doi:10.1073/pnas.1309748110

Showell, C., Binder, O., Conlon, F.L., 2004. T-box genes in early embryogenesis. Dev. Dyn. 229, 201–18. doi:10.1002/dvdy.10480

Struhl, K., 1991. Mechanisms for diversity in gene expression patterns. Neuron 7, 177–81.

Takahashi, H., Hotta, K., Erives, A., Di Gregorio, A., Zeller, R.W., Levine, M., Satoh, N., 1999. Brachyury downstream notochord differentiation in the ascidian embryo. Genes Dev. 13, 1519–1523. doi:10.1101/gad.13.12.1519

Tassy, O., Daian, F., Hudson, C., Bertrand, V., Lemaire, P., 2006. A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. Curr. Biol. 16, 345–58. doi:10.1016/j.cub.2005.12.044

Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R., Laporte, B., Rios, A., Luxardi, G., Kusakabe, T., Joly, J.-S., Darras, S., Christiaen, L., Contensin, M., Auger, H., Lamy, C., Hudson, C., Rothbächer, U., Gilchrist, M.J., Makabe, K.W., Hotta, K., Fujiwara, S., Satoh, N., Satou, Y., Lemaire, P., 2010. The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. Genome Res. 20, 1459–68. doi:10.1101/gr.108175.110

Winchester, J.M., 2016. MorphoTester: An Open Source Application for Morphological Topographic Analysis. PLoS One 11, e0147649. doi:10.1371/journal.pone.0147649

Yagi, K., Takatori, N., Satou, Y., Satoh, N., 2005. Ci-Tbx6b and Ci-Tbx6c are key mediators of the maternal effect gene Ci-macho1 in muscle cell differentiation in Ciona intestinalis embryos. Dev. Biol. 282, 535–549. doi:10.1016/j.ydbio.2005.03.029

Figures

Fig. 1. Data processing. The gene expression information from the ANISEED database is used to produce a 3D model of gene expression. (A) For the 32-cell, 64-cell and 112-cell stages, 3D model embryos from ANISEED were used to generate the 3D model of gene expression for each gene at a cell level. (B) For the tailbud embryos, a 3D model of the tailbud embryo anatomy by Nakamura (Nakamura et al., 2012) is used to produce a 3D model of gene expression at a tissue level.

Fig. 2. Dirichlet normal energy. (A) A schematic tailbud embryo represented by a 3D surface mesh. The area inside the box at the top is shown magnified in (B). (B) The edge vectors (*u* and *v*, black arrows) and normal vectors (red arrows) of a polygon are shown. (C) The end points of the vertex normals form a polygon (with edge vectors n_u and n_v) when translating the normals to a common origin. The DNE measures the spreading of n_u and n_v relative to that of *u* and *v*. Polygons with more curvature will have greater relative spreading of n_u and n_v.

Fig. 3. Relative volume of expression. (A) Distribution plot of the relative volume of expression for all genes in each stage. Red diamonds represent the mean, boxes the interquartile range (IQR). Whiskers the 10 and 90 percentiles. Purple line represents the Max values (the mean of the last decile) and green line the Min values (the mean for the first decile). Development stages on the X-axis. Gray area represents the gastrulation period. The numbers under each bar are the number of genes in each category. Stars represent significant values of pvalues from Kruskal-Wallis test (*<0.05, **<0.01,***<0.001).

(B) Relative volume of transcription factors and non-transcription factors. Black diamonds represent the mean and white circles represent outliers. Black triangles represent the first (lowest) decile of the distribution (10%). Stars represent significance of pvalues from permutation test comparing genes with less or equal than the first decile (*<0.05,**<0.01,***<0.001). 32c, 32-cell stage; 64c, 64-cell stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

Fig. 4. Disparity. (A) Distribution plot of the disparity between regions (cells for early stages and tissues for late stages) of each stage. Distribution plots are represented as in Fig. 2C. (B) Distribution plot of the disparity, considering only transcription factors (white boxes) and non-transcription factors (grey boxes). Stars represent significance of pvalues from permutation test(*<0.05, **<0.01,***<0.001). 32c, 32-cell stage; 64c, 64-cell stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

Fig. 5. DNE through development. Mean DNE values (y-axis) for the six different stages (x-axis) at different scales: (A) with 100 (B) with 1,000 and (C) with 10,000 polygonal faces. As an example, the reconstructed 3D surface (in grey) and the DNE across it (as a heatmap) are shown for

gene expression patterns with low (bottom) and high (top) relative DNE values in the 112-cell and late tailbud stage. Distribution plots are represented as in Fig. 3A. 32c, 32-cell stage; 64c, 64-cell stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud. Stars represent significance of pvalues from one-way ANOVA test (*<0.05, **<0.01,***<0.001)

Fig. 6. Early stages territories. (A) Dendrogram produced by hierarchical clustering, using euclidean distance of the pairwise similarity matrix (pearson's correlation) of all 32-cell, 64-cell and 112-cell stages. The dashed boxes show that the main bifurcation in the tree correspond to the 32-cell and 64-cell stages in one branch and the 112-cell stage (gastrula) on the other. The red arrow shows the cutoff to produce 24 clusters. (B) The names of individual cells, following the nomenclature of Conklin, is indicated with a prefix shown at right. Territories in the 32-cell, 64-cell and 112-cell stages (top, middle and bottom, respectively). Color refers to which synexpression territory of the dendrogram (A) each cell is part of. Animal view based on Nicol and Meinertzhagen, 1988b and vegetal view based on Cole and Meinertzhagen, 2004. In the 112-cell stage, the cell marked with a star (*) is the A7.6 cell, but in our analysis represents their descendant cells (A8.11 and A8.12).

Fig. 7. Tailbud stages territories. A) Dendrogram produced by hierarchical clustering, using a euclidean distance of the pairwise similarity matrix (pearson's correlation) of the early, mid and late tailbud stages. The colored boxes show the cutoff to produce 10 clusters. (B) Territories in the tailbud stages shown in a lateral, para-sagital and sagital views of a 3D embryo model. Color refers to which synexpression territory of the dendrogram (A) each tissue is part of.

Fig. 8. Gene expression similarity between lineage related cells. A) At the left, diagram showing the number of genes expressed differently between each cell and its mother and sister cells (this is genes that are expressed in one cell but not in the other). The cells are represented as filled circles, with color representing its cell fate (color code at the top right; based on Imai, et al., 2006). Circle size represents the number of cell pairs that have exactly the same gene expression pattern. Notice the case of converging lines, that represent two groups of cells with different gene expression in one stage that give rise to identical cells in the next one (i.e., b8.21-32). The width and color of the lines represent the number of genes differentially expressed between cells (color code at the bottom right). For this analysis only genes with expression information in the three early stages and with restricted expression in at least one of them were taken into account (n= 169). B) On top, number of blastomere pairs with a unique expression profile in the early stages (this is a combination of expressed genes not found in any other cells in the embryo). At bottom, the number of blastomere pairs with unique expression profile normalized by the number of blastomere pairs. 32c, 32-cell stage; 64c, 64-cell stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late

tailbud.

SUPPLEMENTARY DATA

S1 Fig. Comparison of the relative volume measure to the proportion of cells with expression. In the top panels, the relative volume of expression is shown for the early stages (left) and the tailbud stages (right). In the bottom panels, the proportion of cells with expression is shown (if a gene has expression in 16 cells at the 32-cell stage its value would be 0.5) for the early stages (left) and tailbud stages (right). 32c, 32-cells stage; 64c, 64-cells stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S2 Fig. On the top panel, distribution plot of the relative volume of transcription factors and non-transcription factors, analyzing separately maternal and zygotic genes (left and right respectively). On the bottom panel, the same for signaling molecules and non-signalling molecules genes. Black diamonds represent the mean and white circles represent outliers. Stars represent significance of pvalues from permutation test (*<0.05,**<0.01,***<0.001). s32c, 32-cell stage; s64c, 64-cell stage; s112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S3 Fig. Relative volume of the transcription factor families. Each plot shows the distribution plot of the relative volume of the genes of each TF family (gray boxes) and the same for the genes that are not part of such family (white boxes). The TF family is shown in the top right of each plot. Black diamonds represent the mean and white circles represent outliers. Stars represent significance of pvalues from permutation test (*<0.05, **<0.01,***<0.001). s32c, 32-cell stage; s64c, 64-cell stage; s112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S4 Fig. Distribution plot of the relative volume of transcription factors of the “Other-TFs” family and the rest of the genes, analyzing separately maternal and zygotic genes (left and right respectively). Black diamonds represent the mean and white circles represent outliers. Stars represent significance of pvalues from permutation test (*<0.05,**<0.01,***<0.001). s32c, 32-cell stage; s64c, 64-cell stage; s112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S5 Fig. Distribution plot of the relative volume of expression for all genes in each stage, excluding transcription factor genes. Red diamonds represent the mean, boxes the interquartile range (IQR). Whiskers the 10 and 90 percentiles. Purple line represents the Max values (the mean of the last decile) and green line the Min values (the mean for the first decile). Development stages on the X-axis. Gray area represents the gastrulation period. Stars represent significant values of pvalues from Kruskal-Wallis test (*<0.05, **<0.01,***<0.001).32c, 32-cell stage; 64c, 64-cell stage; 112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S6 Fig. Disparity of the transcription factor families. Each plot shows the distribution plot of the disparity, considering only a transcription factor family (gray boxes) and the same for the genes that are not part of such family (white boxes). The TF family is shown in the top right of each plot. Stars represent significance of pvalues from permutation test (*<0.05, **<0.01,***<0.001). s32c, 32-cell stage; s64c, 64-cell stage; s112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S7 Fig. Independence of volume of expression and disparity values. As an example, an embryo of six cells (top center) is shown expressing four different gene expression combinations of four genes (A, B, C, D). In the first row, each gene expression configuration is represented as an Euler diagram, with the expression of each gene as a closed curved line, representing the subset of the cells in which it is expressed (in a color code shown at the top left). The same is represented in the third row as binary expression matrix (1 with expression, 0 without expression). In the second row, the mean relative volume of each configuration is shown (0.5). In the fourth row, the pairwise distance between each cell is shown as a matrix. The distance is calculated as 1-(pearsons correlation), so it ranges from 0 to 2. In the final row, distribution plot of the pairwise distances (fourth row) and the mean disparity is shown below in parenthesis.

S8 Fig. Relationship between the roughness (DNE) measure and the relative volume of expression in the 112-cell stage. Each circle in the point represents one gene. The black line represents the regression line. The squared r and pvalue of the linear model are shown. As an example, two genes with similar relative volume but high (green circle) and low (red circle) DNE values are shown at the top and bottom of the scatter plot, respectively.

S9 Fig. Territories and fate map comparison. (A) The dendrogram shows the cell-lineage of the *Ciona intestinalis* embryo from the 32-cell, 64-cell and 112-cell stages (Imai et al., 2004). As the embryos are bilaterally symmetrical, only cells from one half of the embryo are represented. Each cell label is inside a color box that represents the synexpression territory it belongs to in our analysis (as in Fig 6A). At the bottom, schematic embryos of each stage in a vegetal and animal view. Color of each cells refers to which synexpression territory of the each cell is part of. (B) Close to the terminal branches of the cell-lineage dendrogram, the cell fate of each cell is shown as a small as a circle with the color code represented at the right side. The colors of each column in (A) and in (B) are independent.

S10 Fig. As figure 8 but showing the proportion of genes differentially expressed between mother and daughter and sister pair cells (instead of the absolute number of genes differentially expressed

as in Fig. 8).

S11 Fig. Schematic diagram showing all the tissues analyzed in the tailbud stages. Each tissue is shown in a random color.

S1 Table. Relative volume values (in percentage) for all the genes analyzed in this study. Each row is one gene, the first column have the Kyoto Hoya (KH) gene name and the rest of the columns contain the relative volume values for each stage of that gene. NA values mean that there is no available expression information for this gene in the ANISEED database. s32c, 32-cell stage; s64c, 64-cell stage; s112c, 112-cell stage; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud.

S2 Table. List transcription factor genes used in this analysis (from http://ghost.zool.kyoto-u.ac.jp/TF_KH.html).

S3 Table. List of genes differentially expressed between sister cells, as shown in Fig. 8.

FIGURES

Figure 1

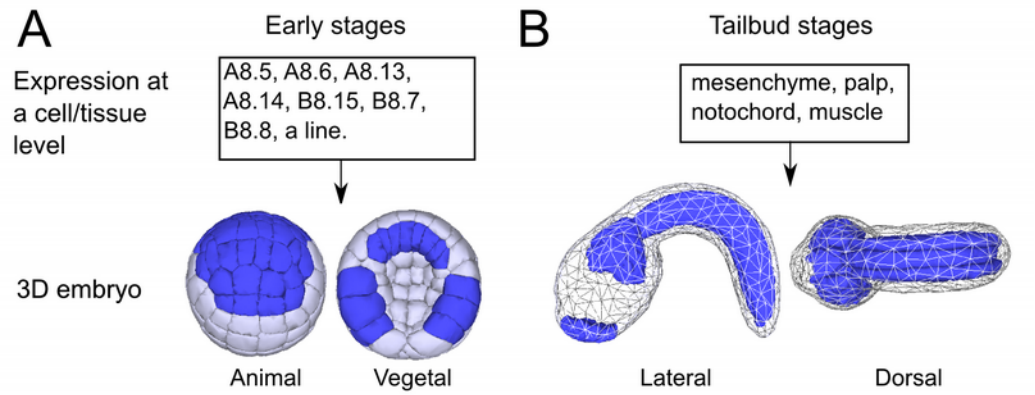

Figure 2

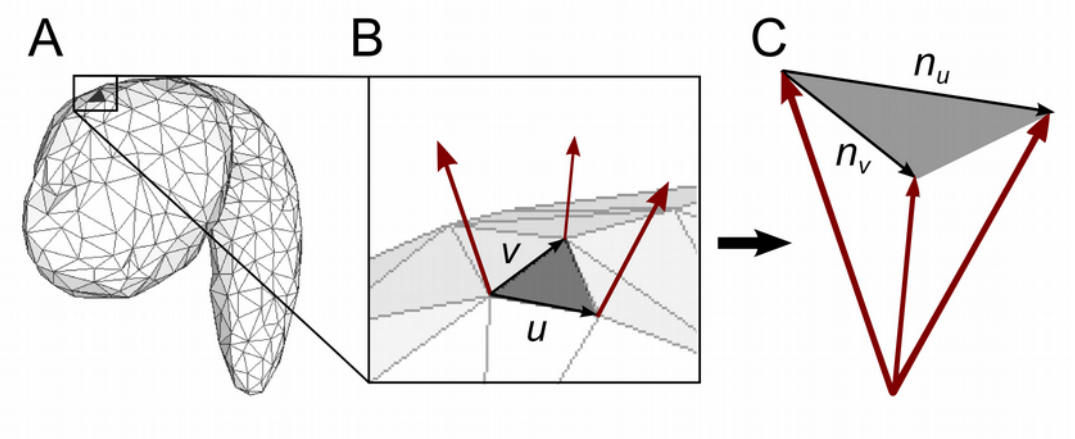

Figure 3

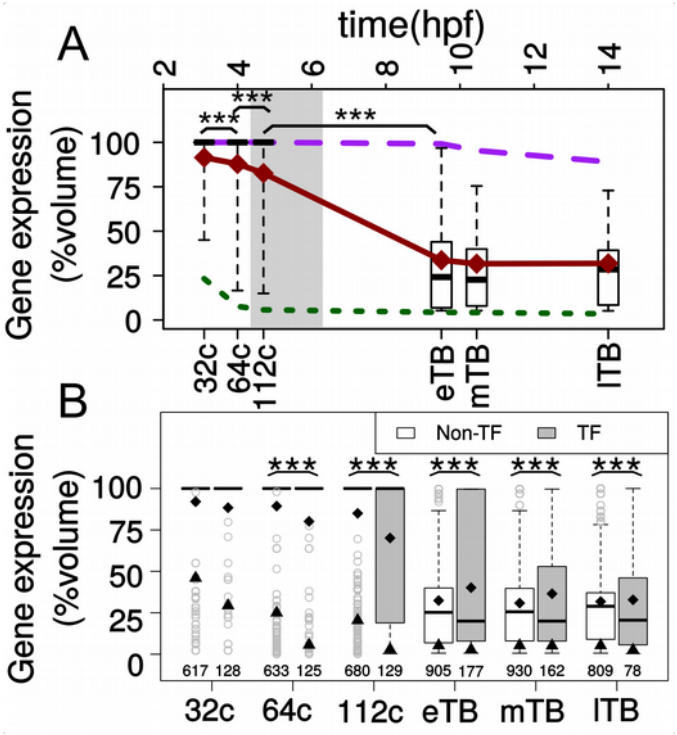

Figure 4

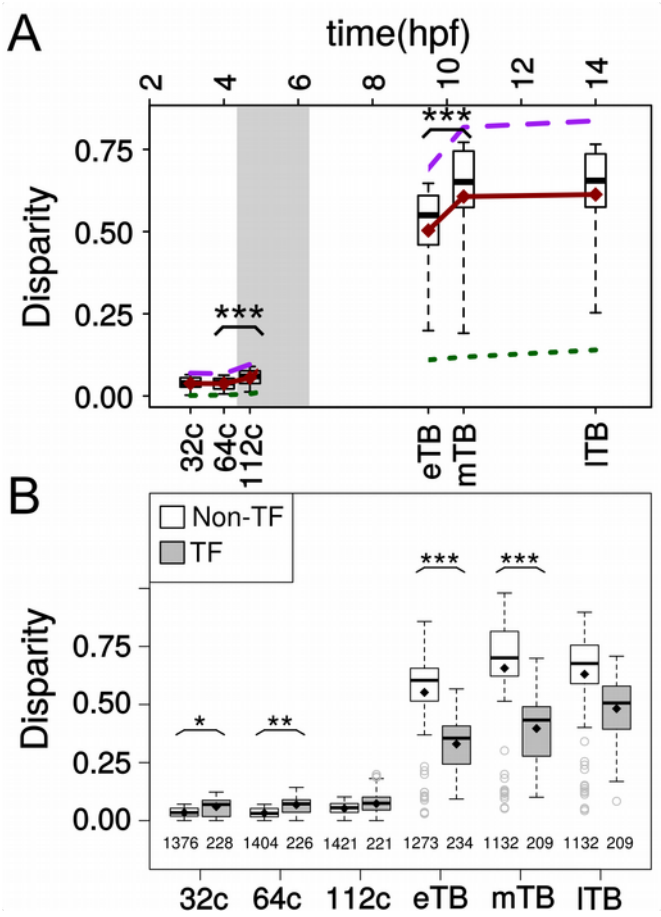

Figure 5

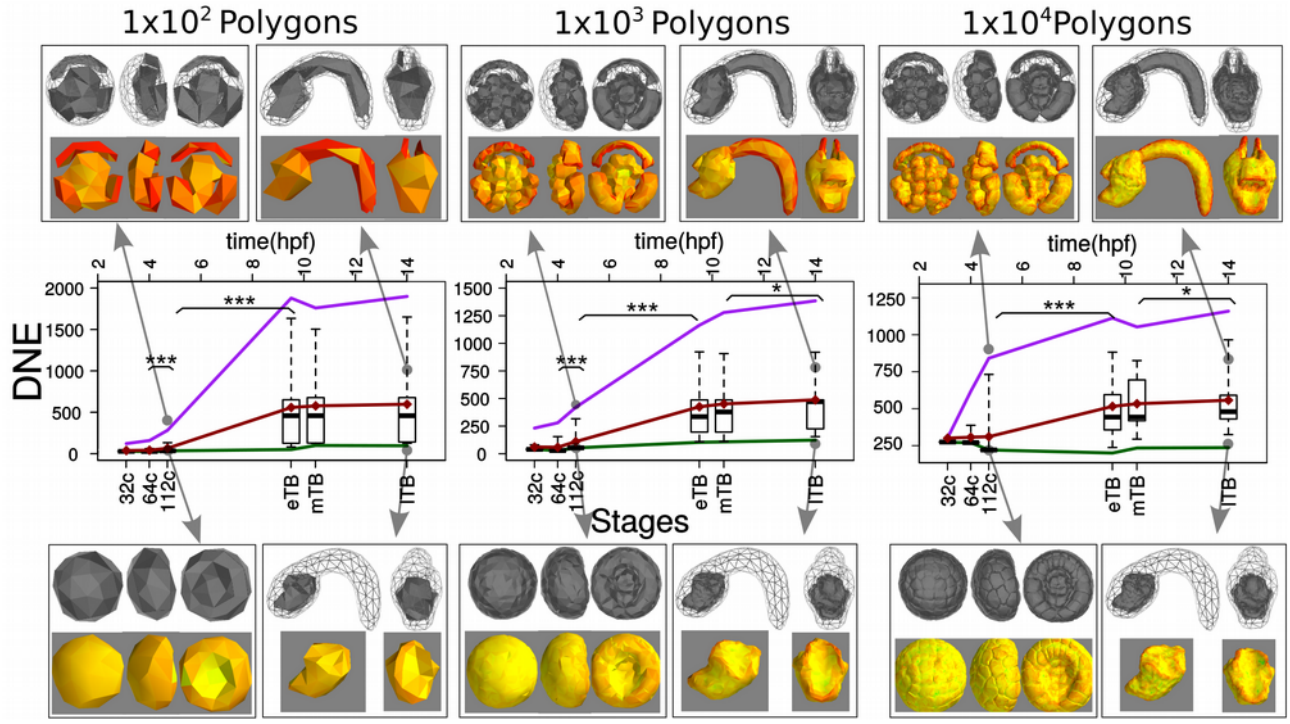

Figure 6

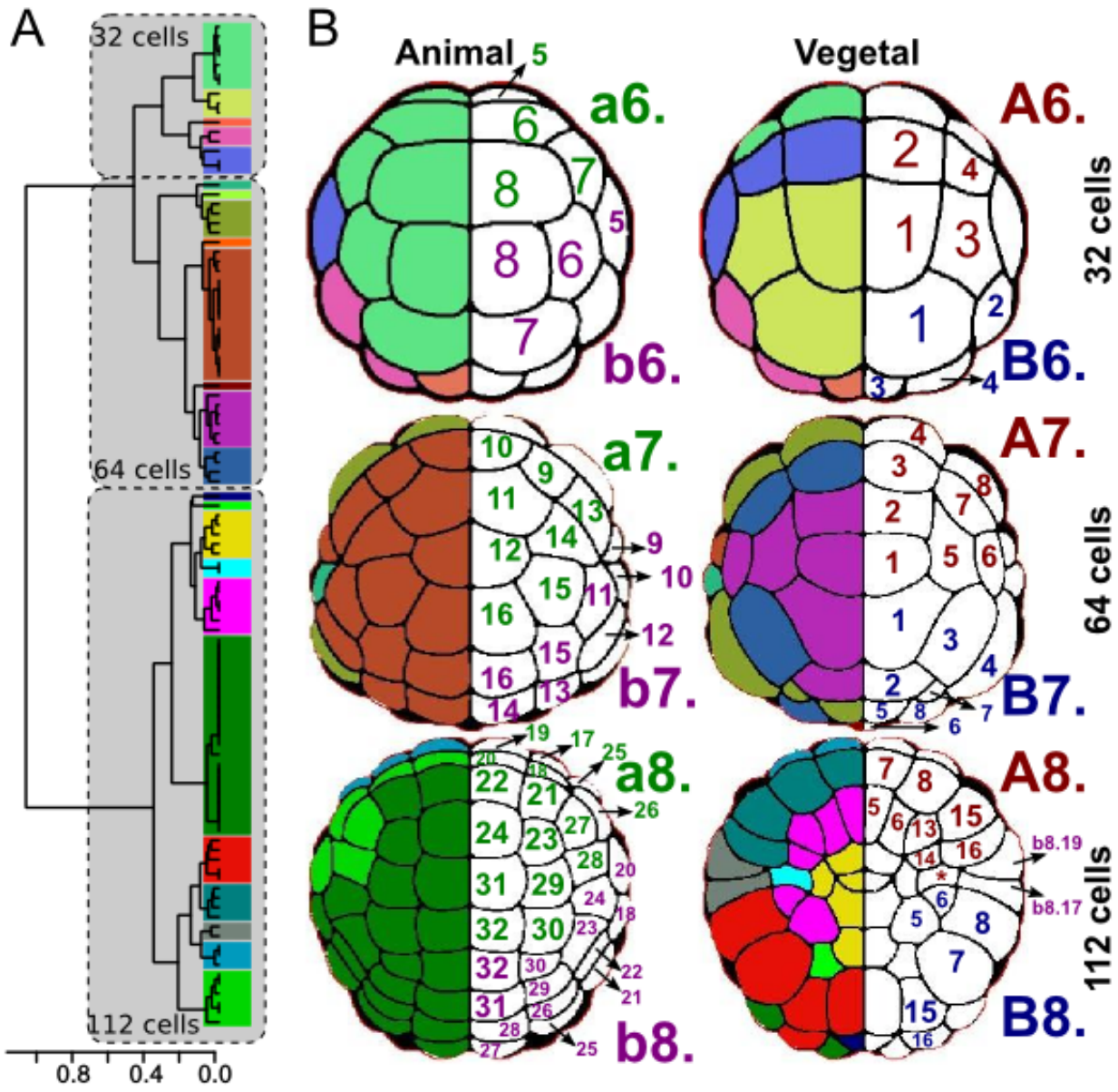
A
32 cells
64 cells
112 cells
0.8
0.4
0.0
B
Animal
Vegetal
a6.
b6.
A6.
B6.
32 cells
a7.
b7.
A7.
B7.
64 cells
a8.
b8.
A8.
B8.
112 cells

Figure 7

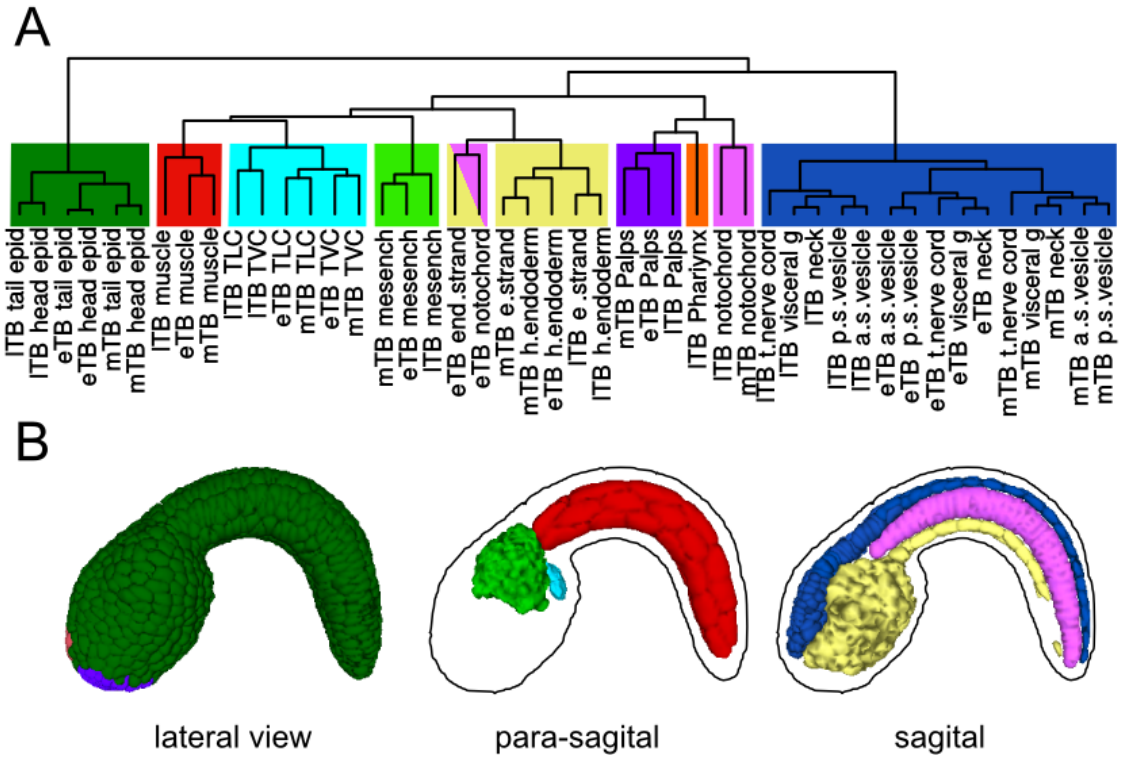

Figure 8

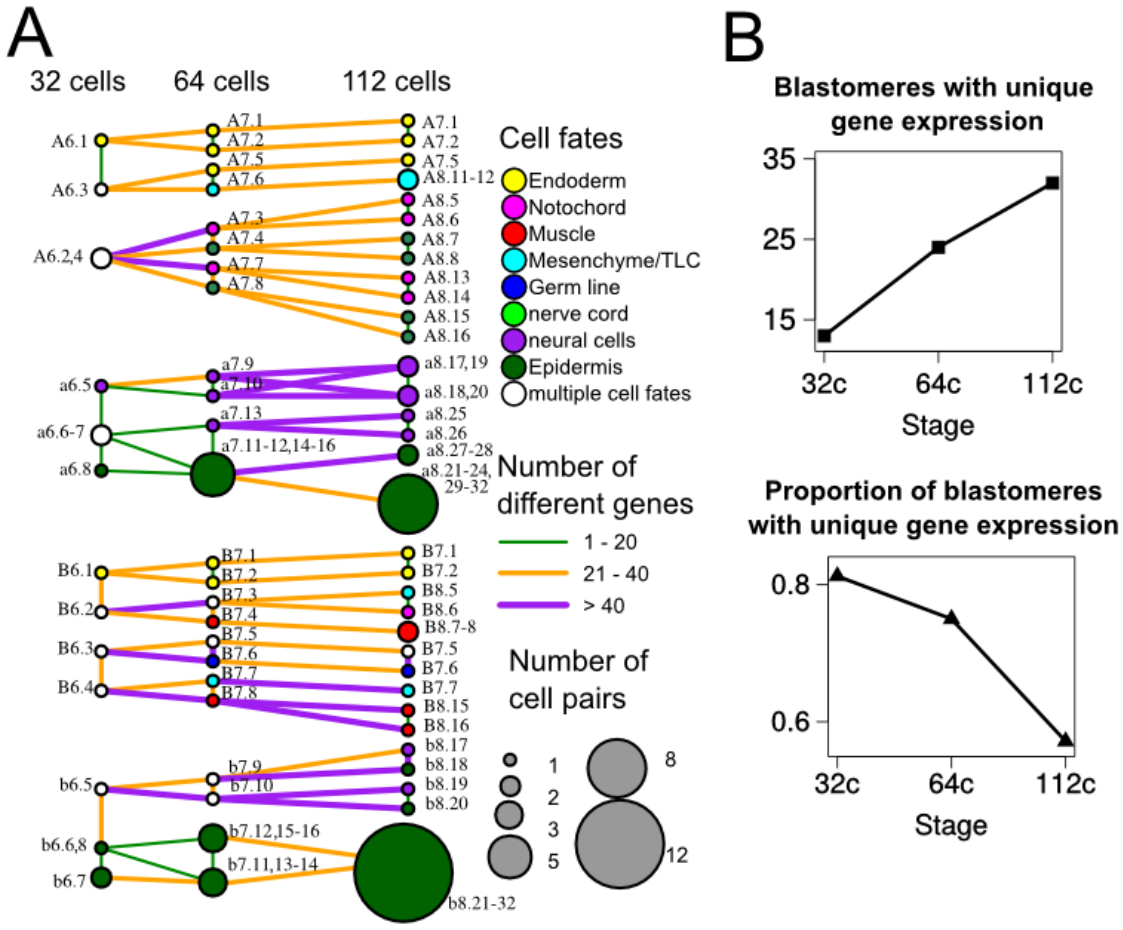

Table 1. List of tissues analyzed in the tailbud stages.

Early tailbud	Mid tailbud	Late tailbud
Head epidermis	Head epidermis	Head epidermis
Tail epidermis	Tail epidermis	Tail epidermis
TLC (Trunk lateral cells)	TLC (Trunk lateral cells)	TLC (Trunk lateral cells)
TVC (Trunk ventral cells)	TVC (Trunk ventral cells)	TVC (Trunk ventral cells)
Mesenchyme	Mesenchyme	Mesenchyme
Head endoderm	Head endoderm	Head endoderm
Notochord	Notochord	Notochord
Endodermal strand	Endodermal strand	Endodermal strand
Palps	Palps	Palps
Tail nerve cord	Tail nerve cord	Tail nerve cord
Anterior sensory vesicle	Anterior sensory vesicle	Anterior sensory vesicle
Posterior sensory vesicle	Posterior sensory vesicle	Posterior sensory vesicle
Visceral glangion	Visceral glangion	Visceral glangion
Neck	Neck	Neck
		Phariynx

Supplementary Figures

Figure S1

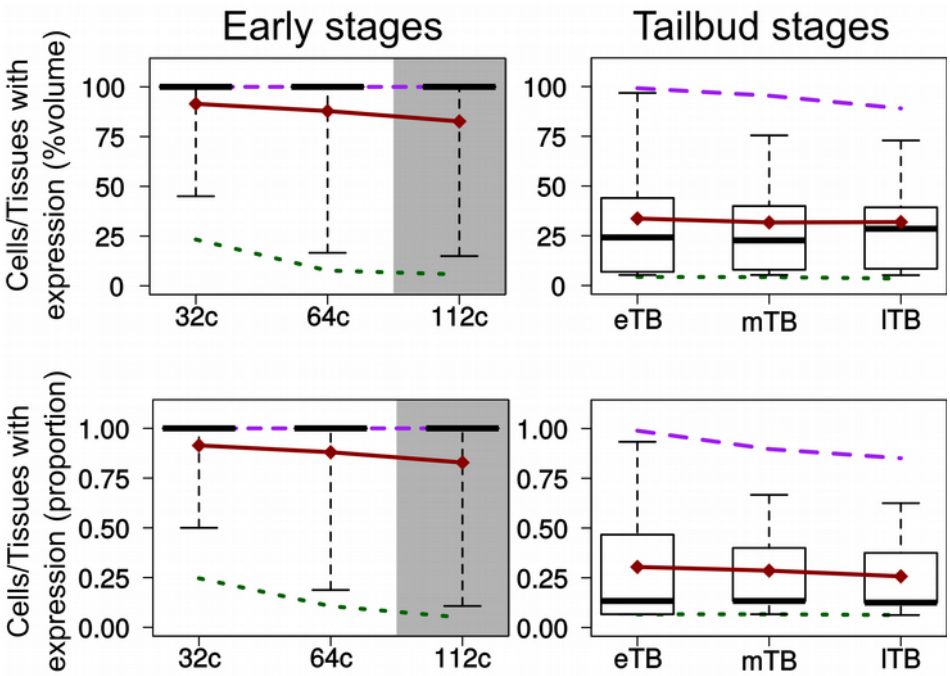

Figure S2

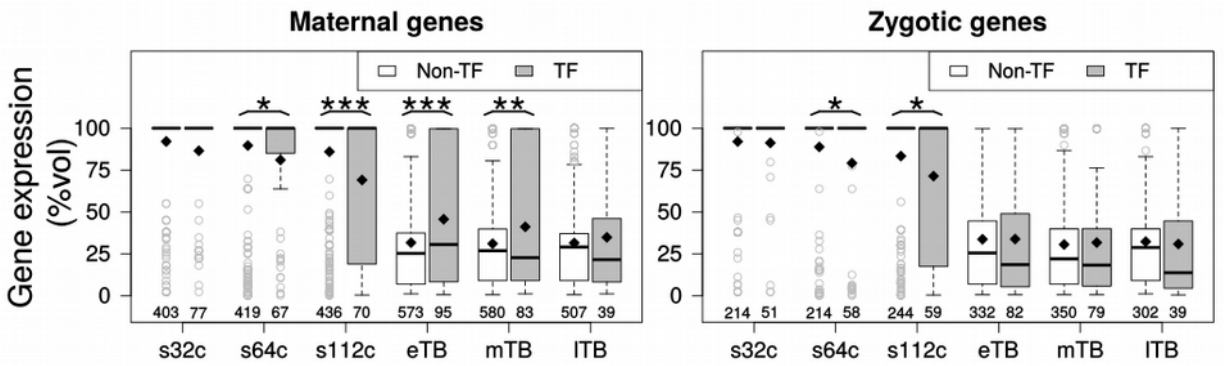

Figure S3

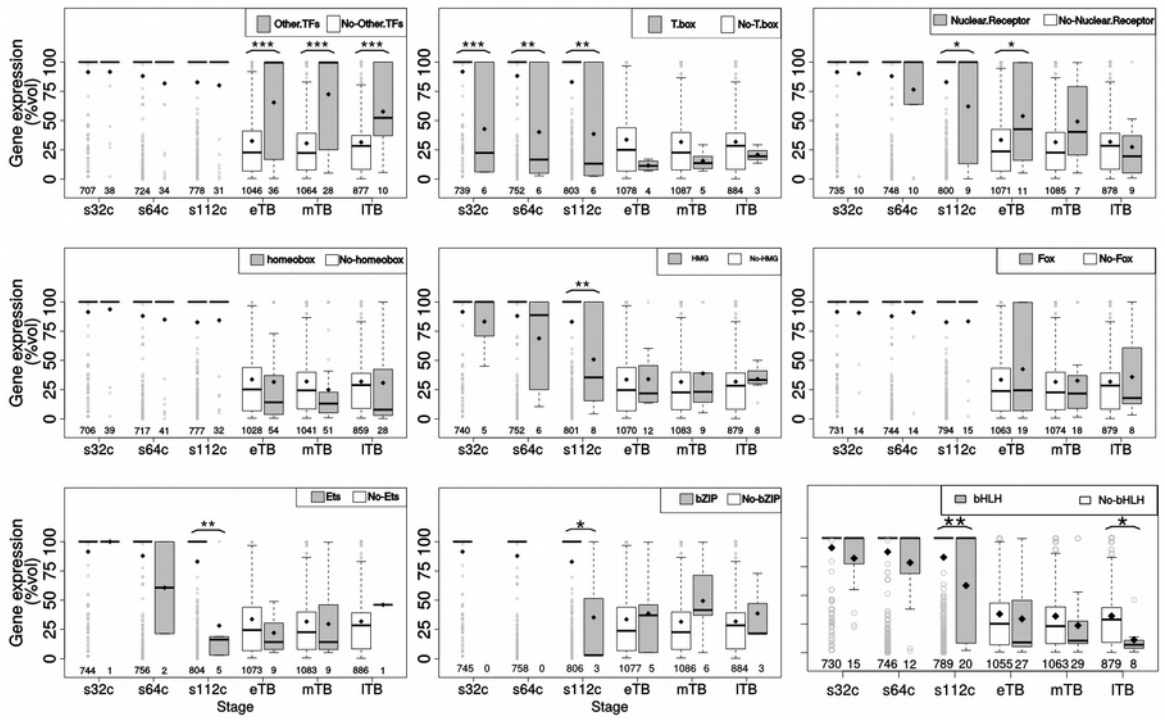

Figure S4

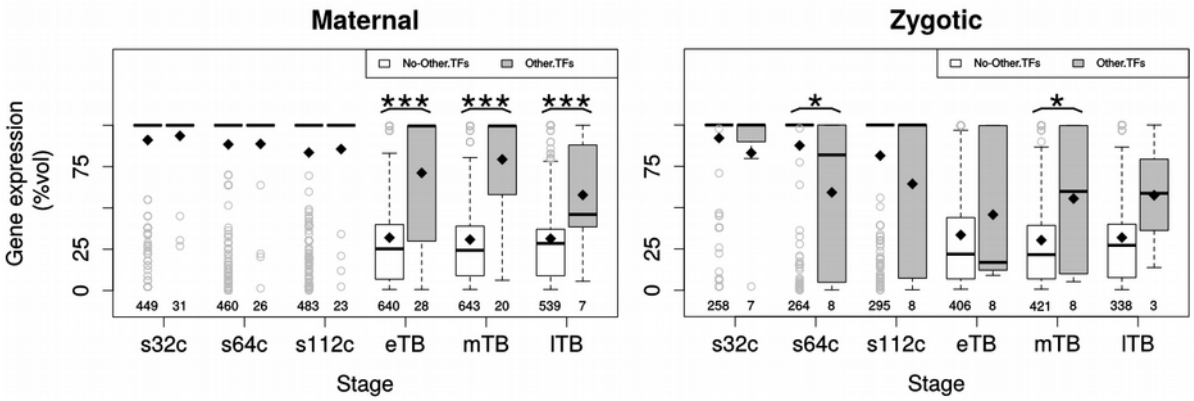

Figure S5

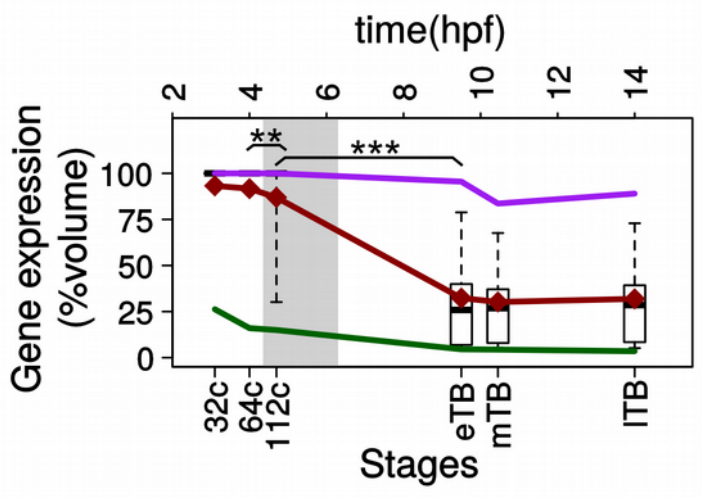

Figure S6

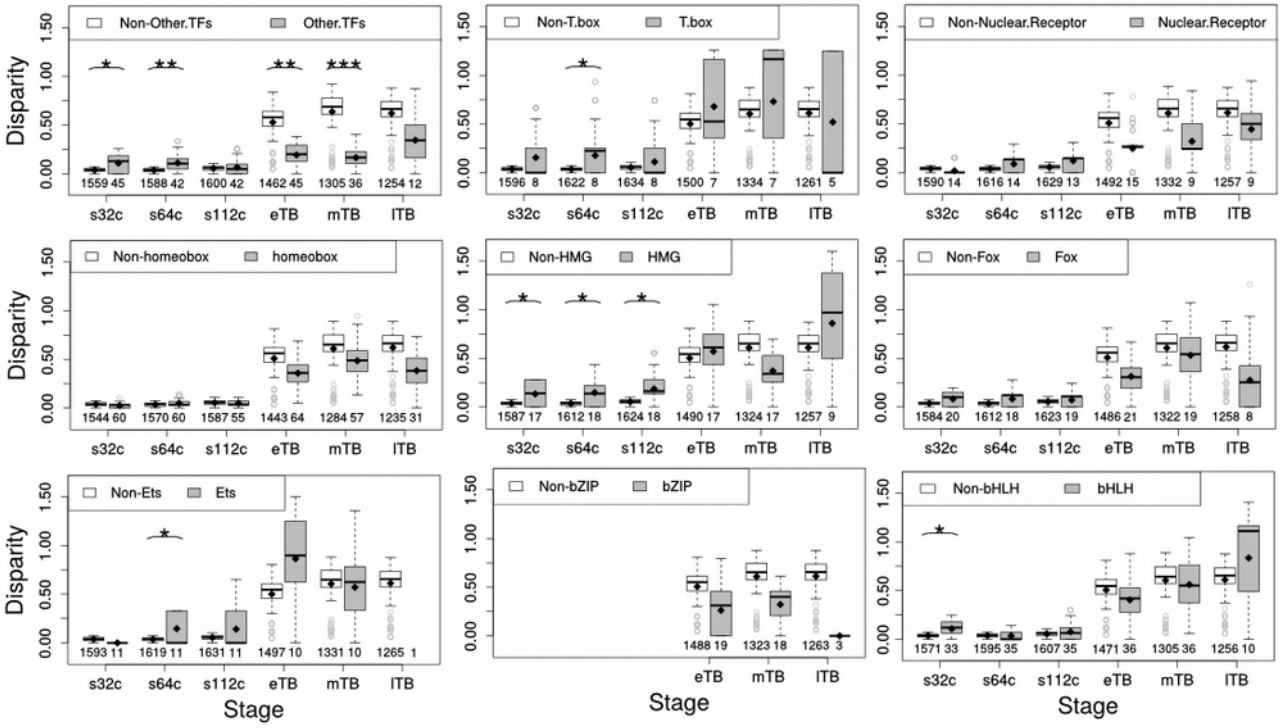

Figure S7

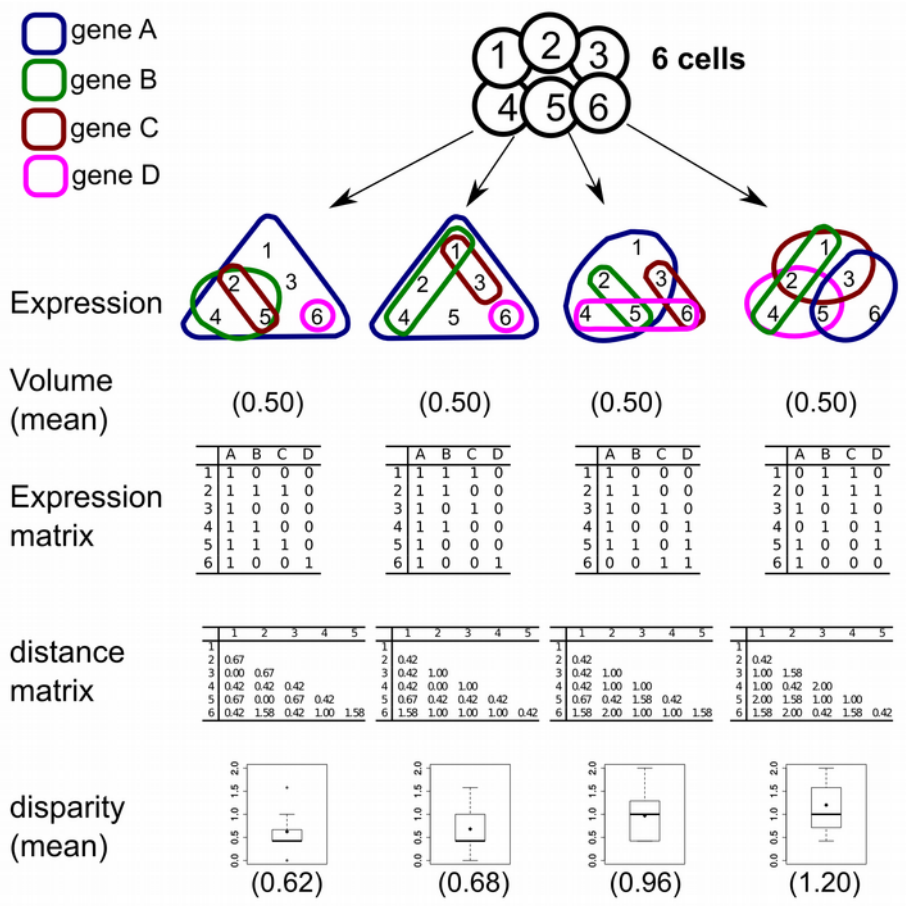

Figure S8

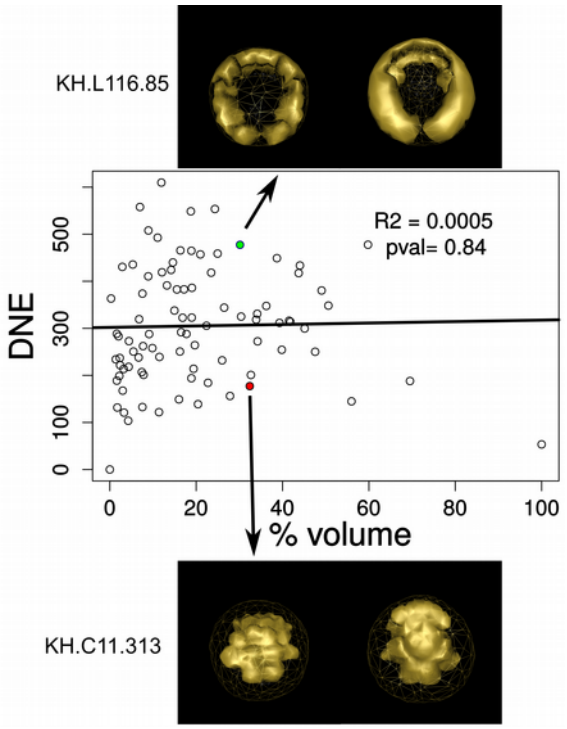

Figure S9

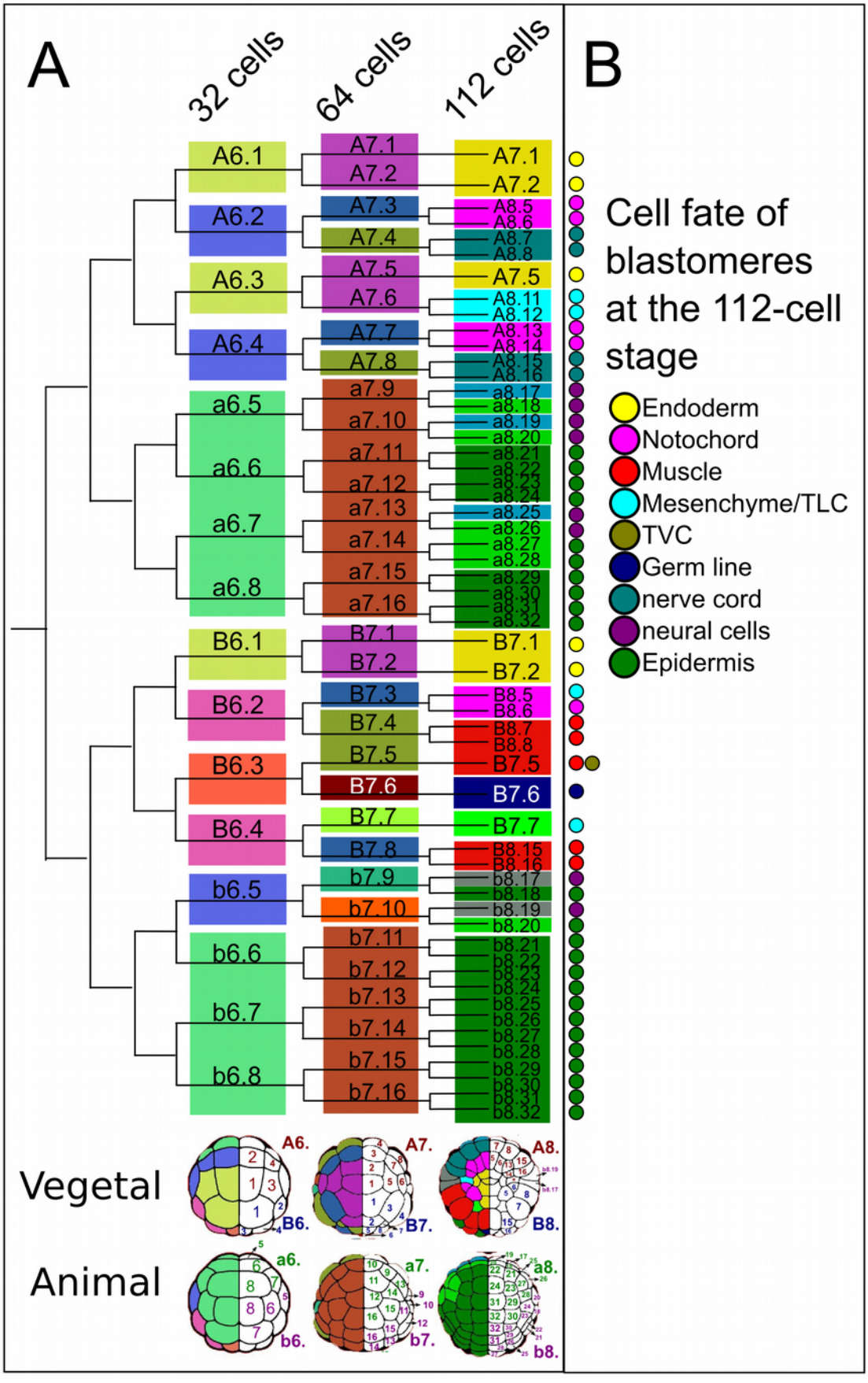
A
32 cells
64 cells
112 cells
A6.1
A6.2
A6.3
A6.4
a6.5
a6.6
a6.7
a6.8
B6.1
B6.2
B6.3
B6.4
b6.5
b6.6
b6.7
b6.8
A7.1
A7.2
A7.3
A7.4
A7.5
A7.6
A7.7
A7.8
a7.9
a7.10
a7.11
a7.12
a7.13
a7.14
a7.15
a7.16
B7.1
B7.2
B7.3
B7.4
B7.5
B7.6
B7.7
B7.8
b7.9
b7.10
b7.11
b7.12
b7.13
b7.14
b7.15
b7.16
A7.1
A7.2
A8.5
A8.6
A8.7
A8.8
A7.5
A8.11
A8.12
A8.13
A8.14
A8.15
A8.16
a8.17
a8.18
a8.19
a8.20
a8.21
a8.22
a8.23
a8.24
a8.25
a8.26
a8.27
a8.28
a8.29
a8.30
a8.31
a8.32
B7.1
B7.2
B8.5
B8.6
B8.7
B8.8
B7.5
B7.6
B7.7
B8.15
B8.16
b8.17
b8.18
b8.19
b8.20
b8.21
b8.22
b8.23
b8.24
b8.25
b8.26
b8.27
b8.28
b8.29
b8.30
b8.31
b8.32
B
Cell fate of blastomeres at the 112-cell stage
Endoderm
Notochord
Muscle
Mesenchyme/TLC
TVC
Germ line
nerve cord
neural cells
Epidermis
Vegetal
Animal
A6.
A7.
A8.
B6.
B7.
B8.
a6.
a7.
a8.
b6.
b7.
b8.

Figure S10

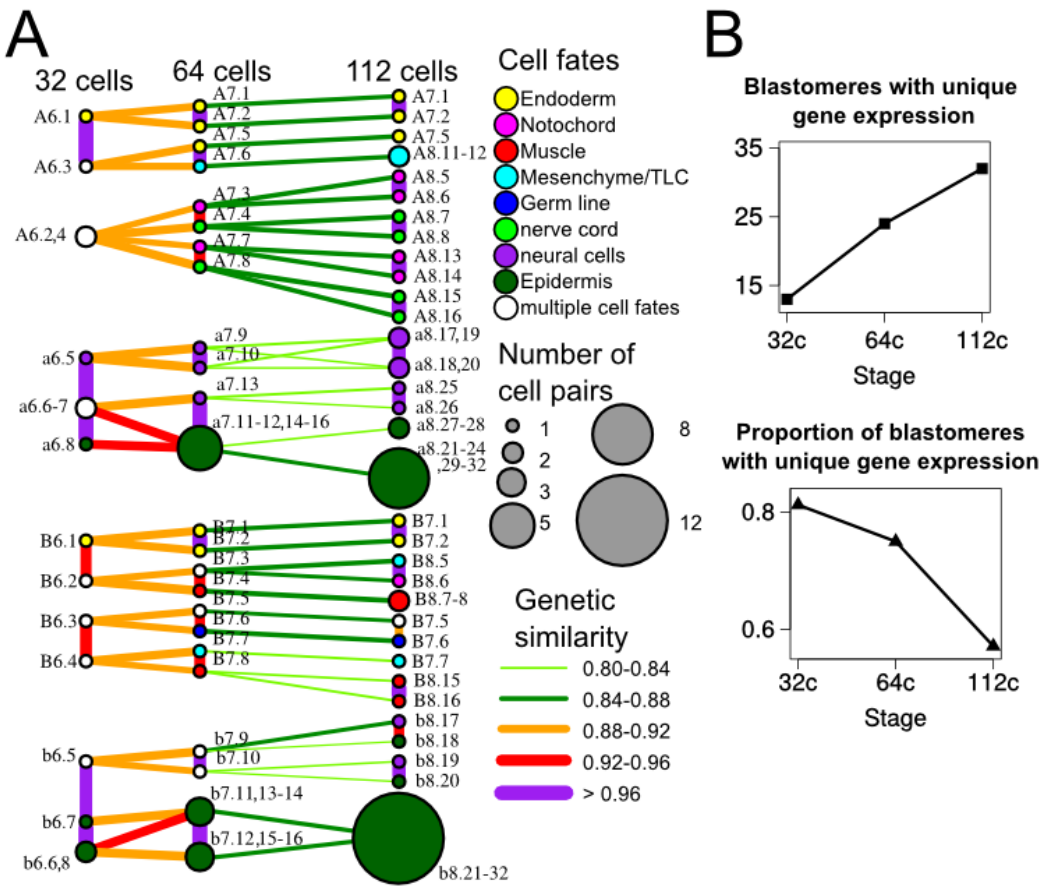

Figure S11